

Optimasi Parameter Random Forest Menggunakan Grid Search Untuk Analisis Time Series

Dzikri Ziaul Haq Iskandar ¹⁾*, Yudi Ramdhani ²⁾

Teknik Informatika, Universitas Adhirajasa Reswara Sanjaya
iskandarziaulhaq@gmail.com ¹⁾*, yudi@ars.ac.id ²⁾

Abstrak

Dalam membeli dan menjual di bursa saham, prediksi harga saham memainkan peran penting bagi para investor. Namun, prediksi harga saham merupakan tantangan karena dipengaruhi oleh faktor-faktor yang sulit diprediksi, seperti kondisi pasar, kinerja perusahaan, dan berita ekonomi. Oleh karena itu, penelitian ini bertujuan untuk menganalisis deret waktu harga penutupan saham Microsoft Corporation menggunakan algoritma Random Forest (RF) dan mengoptimalkan parameter algoritma dengan menggunakan metode optimisasi grid parameter. Data yang digunakan meliputi rentang waktu dari 1 Maret 1986 hingga 25 Mei 2023 dengan total 9378 catatan dan 6 atribut. Setelah pengumpulan dan pemrosesan data, termasuk verifikasi nilai yang hilang, data dibagi menjadi data pelatihan dan pengujian menggunakan validasi pemisahan. Selain itu, validasi silang digunakan untuk membandingkan algoritma-algoritma dan memilih algoritma RF sebagai model terbaik berdasarkan nilai RMSE terendah. Tingkat optimalisasi parameter dicapai dengan mengoptimalkan parameter grid, dengan Jumlah pohon dan Kedalaman maksimal sebagai parameter yang dioptimalkan. Analisis Paired Two Sample T-Test juga mengungkapkan perbedaan yang signifikan antara RMSE sebelum dan setelah optimisasi. Kesimpulannya, penelitian ini memberikan kontribusi yang signifikan dalam penggunaan algoritma RF dan metode optimisasi parameter grid dalam analisis deret waktu harga penutupan, dengan potensi aplikasi yang signifikan dalam pengambilan keputusan investasi di pasar saham

Kata kunci: Microsoft Corporation, Optimize Parameter Grid, Prediksi, Random Forest, Time Series.

Abstract

[Random Forest Parameter Optimization Using Grid Search for Time Series Analysis]. This study focuses on analyzing the Time Series of Microsoft Corporation stock closing prices and predicting future stock prices using the Random Forest (RF) algorithm. The research aims to address the challenges of stock price prediction due to unpredictable factors like market conditions, company performance, and economic news. The dataset used covers a period from March 1, 1986, to May 25, 2023, comprising 9378 records and 6 attributes. Data preprocessing involved handling missing values and splitting the data into training and testing sets using split validation. Cross-validation was employed to compare different algorithms, with the RF algorithm selected as the best model based on the lowest Root Mean Square Error (RMSE) value. The study further optimized the RF algorithm's parameters, specifically the number of trees and max depth, using the Optimize Parameter Grid method. The optimization process successfully demonstrated a significant improvement in RMSE through a Paired Two Sample T-Test analysis. Overall, this research contributes to the effective use of the RF algorithm and parameter optimization techniques for analyzing Time Series data, with potential applications in supporting investment decisions in the stock market.

Keywords: Microsoft Corporation, Optimize Parameter Grid, Prediksi, Random Forest, Time Series.

1. PENDAHULUAN

Dalam bursa jual beli saham, ada banyak instrumen keuangan. Salah satunya dikenal sebagai instrumen keuangan jangka panjang yang bertukar nilai aset mereka di pasar modal. Memprediksi harga saham merupakan isu yang paling menarik dan kompleks bagi para akademik, investor, dan analis [1]. Harga saham dipengaruhi oleh berbagai faktor, termasuk situasi ekonomi, politik, sosial, dan psikologis, yang sulit dimodelkan secara matematis. Selain itu, sifat harga saham yang dinamis, nonlinear, dan stokastik

menambah kompleksitas dalam melakukan prediksi [2]. Memprediksi harga saham adalah tugas yang menantang dalam analisis keuangan karena volatilitas pasar, ketidakpastian pengaruh eksternal, asimetri informasi, dan tantangan dalam analisis *Time Series*. Kerugian prediksi harga saham dapat berdampak pada kerugian finansial bagi investor. [3]. Microsoft Corporation (MSFT) adalah salah satu perusahaan teknologi terbesar di dunia yang bekerja di bidang perangkat lunak, perangkat keras, layanan *cloud*, dan media sosial [4]. Harga saham MSFT dipengaruhi oleh

berbagai variabel *internal* dan *eksternal*, termasuk kinerja keuangan, inovasi produk, persaingan pasar, suasana hati investor, dan peristiwa dunia [5].

Metode atau algoritma diperlukan untuk mempelajari pola harga saham atau tren dari data historis dan faktor lain dengan baik, dan dapat beradaptasi dengan perubahan atau anomali yang terjadi di pasar saham. Salah satu metode yang paling banyak digunakan untuk memprediksi harga saham adalah *machine learning* [6]. *Machine learning* adalah cabang ilmu komputer yang memungkinkan komputer atau program untuk belajar dari data [7]. *Machine learning* dapat digunakan untuk berbagai tujuan, termasuk prediksi harga saham. *Machine learning* dapat memanfaatkan data historis dan faktor lain yang mempengaruhi pasar saham untuk membedakan pola atau tren dalam harga saham dan menggunakannya untuk memprediksi harga saham di masa depan [8]. *Machine learning* dan analisis *Time Series* adalah metode dominan dalam prediksi harga saham. Analisis *Time Series* merupakan teknik statistik yang mempelajari perilaku interval waktu fluktuatif untuk mengidentifikasi tren, musiman, siklus, dan komponen lainnya. Dengan menggunakan data masa lalu dan saat ini, analisis *Time Series* dapat memprediksi nilai masa depan dari variabel tersebut. Beberapa model umum yang digunakan dalam analisis *Time Series* adalah *Autoregressive Integrated Moving Average* (ARIMA), *seasonal ARIMA*, *Autoregressive Moving Average* (ARMA), dan *eksponensial smoothing* [9]. Model analisis *Time Series* klasik memiliki batasan seperti asumsi linearitas, stagnasi, normalitas, dan independensi antar pengamatan. Model ini juga rentan terhadap *outlier* dan *noise* dalam data. Untuk mengatasi keterbatasan ini, peneliti telah mengusulkan penggunaan metode *machine learning* dalam memprediksi harga saham. *Machine learning* adalah bidang ilmu komputer yang memungkinkan sistem belajar dari data tanpa pemrograman eksplisit. Dengan menggunakan algoritma yang adaptif dan fleksibel, *machine learning* dapat mengekstrak fitur dan pola yang relevan dari data yang kompleks dan heterogen [10].

Random Forest merupakan metode yang menggunakan serangkaian *decision tree* yang dibangun secara acak dari sejumlah fitur atau variabel input [11]. Algoritma *Random Forest* memiliki beberapa keuntungan dalam memprediksi seri waktu penutupan harga saham MSFT. Pertama, algoritma ini mampu mengelola data besar, kompleks, dan nonlinear dengan baik, sehingga dapat memanfaatkan berbagai fitur atau variabel *input* yang beragam untuk membuat model yang akurat dan kuat [12]. Kedua, algoritma ini mampu mengurangi overfit dan meningkatkan akurasi prediksi dengan melakukan *voting* atau agregasi hasil prediksi dari setiap pohon keputusan, menghasilkan prediksi yang stabil dan konsisten dengan mengurangi varians dan bias [3].

Metode yang digunakan adalah *Support Vector Regression* (SVR) untuk memprediksi harga saham

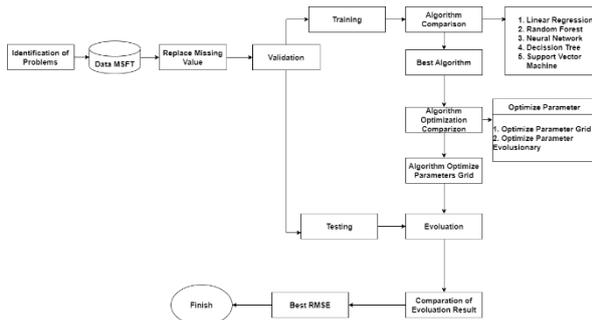
INDF dan MYOR. Algoritma *Grid Search* digunakan untuk menemukan parameter optimal. Hasilnya menunjukkan bahwa metode SVR dengan *Grid Search* dapat menghasilkan prediksi harga saham yang akurat, dengan MAPE sebesar 1,50% dan 2 dari 79,9% untuk INDF, serta MAPE sebesar 2,95% dan 2 dari 96% untuk MYOR [13]. Penggunaan algoritma SVM dengan optimasi parameter menggunakan algoritma genetik dapat meningkatkan prediksi harga saham PT. Antm.JK. Setelah melakukan optimalisasi parameter, hasil prediksi memiliki nilai RMSE terendah sebesar 10.495. Metodologi penelitian yang digunakan meliputi tindakan penelitian, eksperimen, penelitian kasus, dan survei. Temuan penelitian menunjukkan bahwa penerapan algoritma SVM dengan optimasi parameter menggunakan algoritma genetik dapat meningkatkan prediksi harga saham PT. Antm.JK. Selain itu, penelitian ini menekankan bahwa data mining dapat digunakan untuk memprediksi harga saham dengan mendeteksi pola dari sekelompok database yang memberikan informasi penting pada dataset yang besar. Penelitian ini juga melibatkan diagram penutupan harga saham Antm.JK dan diagram perbaikan algoritma prediksi SVM dan SVM+GA. [7].

Penelitian ini bertujuan untuk memprediksi harga penutupan saham Microsoft Corporation (MSFT) menggunakan data historis dalam bentuk *Time Series*. Harga saham MSFT dipengaruhi oleh berbagai faktor dan memiliki volatilitas tinggi, membuat prediksi yang akurat menjadi sulit. Oleh karena itu, pemilihan algoritma yang tepat sangat penting dalam membuat prediksi harga saham yang akurat. Dalam penelitian ini, menggunakan lima algoritma *machine learning*, yaitu *K-Nearest Neighbor* (K-NN), *Neural Network* (NN), *Linear Regression* (LR), *Decision Tree*, dan *Random Forest* (RF). Performa masing-masing algoritma dibandingkan dengan menggunakan metrik *Root Mean Squared Error* (RMSE), yang mengukur tingkat penyimpangan antara nilai aktual dan prediksi. Penelitian ini juga menyelidiki pengaruh fitur input dan parameter pada akurasi prediksi. Diharapkan bahwa penelitian ini dapat memberikan pemahaman yang lebih baik tentang kemampuan algoritma-algoritma tersebut dalam menganalisis *Time Series* dan memprediksi harga penutupan saham MSFT. Hasilnya dapat bermanfaat bagi investor dan peserta pasar modal dalam membuat keputusan investasi yang lebih baik berdasarkan prediksi yang lebih akurat.

2. BAHAN DAN METODE

2.1. Desain Penelitian

Optimize Parameter Grid digunakan dalam penelitian ini untuk memilih parameter terbaik, dan metode *Random Forest* (RF) menggunakan beberapa tahapan dalam metode penelitian untuk melakukan prediksi penutupan harga saham MSFT. Desain tahapan penelitian yang digunakan dapat dilihat pada Gambar 1.



Gambar 1 Desain Penelitian

2.1.1. Identifikasi Masalah

Karena sejumlah keterbatasan, sulit untuk membuat prediksi yang tepat dari data *Time Series* saham MSFT. Algoritma *Random Forest* digunakan sebagai pendekatan prediksi untuk memecahkan masalah ini. Untuk meningkatkan keefektifan algoritma *Random Forest* dalam memprediksi harga saham MSFT dari waktu ke waktu, parameter yang tepat harus dimodifikasi. Parameter optimal untuk memaksimalkan estimasi *Time Series* harga saham MSFT menggunakan teknik *Random Forest* diantisipasi untuk diidentifikasi melalui optimasi parameter.

2.1.2. Dataset

Untuk mengumpulkan data, peneliti memanfaatkan dataset publik yang tersedia di Yahoo Finance. Data tersebut bersifat *Time Series* dan mencakup rentang waktu mulai dari 1 Maret 1986 hingga 25 Mei 2023. Setelah dilakukan pengumpulan, dataset tersebut terdiri dari 9378 record data dengan 6 atribut. Penjelasan atribut ditunjukkan pada tabel 1.

Tabel 1 Penjelasan Atribut

Data Input	Rincian
Date	Tanggal transaksi saham MSFT
Open	Harga saham MSFT pada saat pasar dibuka
High	Harga tertinggi saham MSFT yang dicapai dalam satu hari
Low	Harga terendah saham MSFT yang dicapai dalam satu hari
Close	Harga saham MSFT pada saat pasar ditutup
Volume	Jumlah saham MSFT yang diperdagangkan dalam satu hari

2.1.3. Replace Missing Value

Peneliti melakukan *preprocessing* data dengan mengecek dataset menggunakan operator *filter example* untuk mengatasi *missing value*. *Missing value* adalah nilai yang hilang pada data yang seharusnya ada. Hal ini bisa disebabkan oleh kesalahan input, kerusakan file, atau ketidaktahuan responden. *Missing value* dapat mengganggu analisis data dan mereduksi validitas hasil penelitian. Oleh karena itu, peneliti perlu melakukan pengecekan *missing value* dan mengambil tindakan yang tepat, seperti

menghapus baris atau kolom yang mengandung *missing value*, mengisi *missing value* dengan rata-rata atau *median*, atau menggunakan metode imputasi yang lebih canggih.

2.1.4. Validasi Data

Langkah validasi data dilakukan oleh penulis dengan membagi data penelitian menjadi data *training* dan *testing*. Untuk memisahkan data, para peneliti menggunakan *cross validation* dan *split validation*. *Cross validation* digunakan untuk mengidentifikasi kinerja optimal dari model yang dipertimbangkan, sedangkan *Split validation* adalah pendekatan validasi yang secara acak membagi data dengan perbandingan spesifik antara data *training* dan *testing* dengan rasio 0,9 sampai dengan 0,5 [14].

2.1.5. Komparasi Algoritma

Pada tahap perbandingan algoritma, dilakukan *testing* menggunakan beberapa algoritma yang berbeda. Penelitian ini melibatkan 5 algoritma yang terdiri dari *Linear Regression* (LR), *Neural Network* (NN), *Decision Tree*, *K-Nearest Neighbor* (K-NN) dan *Random Forest*.

a. Linear Regression

Linear regression merupakan teknik statistik yang digunakan untuk menetapkan model matematika yang menggambarkan hubungan antara dua variabel dengan memasukkan garis lurus. Teknik regresi linear digunakan untuk membuat prediksi tentang nilai variabel terikat (Y) dengan mempertimbangkan nilai variabel bebas. (X). Regresi linear dapat digunakan untuk mengidentifikasi hubungan antara variabel bebas variabel terikat dalam penelitian kuantitatif [15].

b. Neural Network

Neural network adalah kumpulan unit *input/output* yang terhubung dengan bobot masing-masing. Dalam fase penyesuaian, bobot *neural network* diubah agar dapat memprediksi kelas yang tepat dari tupel [16]. Setiap *neuron* menerima input dengan bobot masukan yang ditentukan, kemudian dievaluasi menggunakan fungsi propagasi yang menjumlahkan nilai bobot masukan. Hasilnya dibandingkan dengan ambang rata-rata untuk aktivasi *neuron*. Pembelajaran *backpropagation* melibatkan perubahan bobot *neuron* ke belakang berdasarkan nilai kesalahan. Untuk mendapatkan kesalahan, langkah penyebaran maju harus diselesaikan terlebih dahulu. *Neuron* dalam lapisan tersembunyi diaktivasi menggunakan fungsi aktivasi seperti fungsi sigmoid [17].

c. Decision Tree

Algoritma *Decision Tree* secara umum dan sangat efektif untuk membuat prediksi dan klasifikasi yang akurat. Algoritma *decision tree* memiliki kemampuan untuk mengubah informasi fakta menjadi *decision tree* terstruktur yang secara efektif mewakili sebuah aturan dan aturan tersebut mudah untuk dimengerti. Proses yang dilakukan pada metode *DecisionTree* melibatkan perubahan data menjadi struktur pohon keputusan, mengubah struktur pohon menjadi format aturan dan kemudian menyederhanakan aturan [18], [19].

d. K-Nearest Neighbor

K-Nearest Neighbor (K-NN) adalah metode pembelajaran berbasis pengawasan yang digunakan untuk klasifikasi objek baru berdasarkan sampel latihan. Metode ini mengandalkan jumlah tetangga terdekat dari titik uji untuk menentukan kelasnya. Klasifikasi dilakukan dengan memilih kelas yang paling banyak muncul di antara tetangga terdekat. Algoritma K-NN tidak memerlukan model yang disesuaikan dan bergantung pada jarak *Euclidean* untuk menentukan tetangga terdekat [20].

e. Random Forest

Tahap ini menunjukkan bahwa algoritma *Random Forest* adalah model terbaik untuk prediksi penutupan harga saham MSFT. *Random Forest* dipilih berdasarkan nilai RMSE terendah dibandingkan dengan algoritma lainnya. Namun, *Random Forest* menghadapi masalah dalam menentukan atribut dan ambang batas yang optimal. Untuk mengatasi masalah ini, digunakan dua metode ekstraksi fitur: *entropy* dan *information gain*. Metode *entropy* digunakan untuk mengukur ketidakpastian dan mencari atribut dan ambang batas yang menghasilkan ketidakpastian terendah. Sedangkan metode *information gain* digunakan untuk menghitung seberapa banyak atribut memberikan penjelasan tentang kelasnya.[8].

$$Entropy = - \sum_{i=1}^n p_i \log_2 p_i$$

$$Information\ Gain = EP - EA$$

Keterangan:

n = Jumlah Kelas

p_i = Probabilitas Kemunculan kelas ke i

EP = Nilai *Entropy* dan *Bootstrapped dataset*

EA = Nilai *Entropy* dari atribut yang dipilih.

Tahapan dari algoritma *Random Forest* terdiri dari beberapa langkah. Pertama, dibuat dataset *bootstrapped* dengan cara secara acak memilih sampel data sebanyak jumlah data yang digunakan dari data *training* dengan penggantian (*replacement*). Selanjutnya, fitur-fitur dari dataset *bootstrapped* dipilih secara acak tanpa penggantian untuk menghitung nilai *entropy* dan *information gain*. Langkah berikutnya adalah menggunakan fitur dan *threshold* dengan nilai *information gain* terbaik untuk membentuk aturan pada setiap *node* keputusan (*decision node*) dalam *decision tree* pada setiap tingkat kedalaman. Proses ini diulangi untuk membentuk sejumlah n *decision tree*. Akhirnya, pada tahap evaluasi, data uji diberikan kepada n *decision tree* yang telah dilatih, dan dilakukan *voting* untuk memprediksi label kelas dari data uji tersebut [8].

2.1.6. Komparasi Optimasi Algoritma

Dalam tahap perbandingan algoritma optimasi, penelitian ini menguji tiga fitur optimasi parameter: *Optimize Parameter Grid* dan *Optimize Parameter Evolutionary*. *Optimize Parameter Evolutionary* terbukti lebih akurat daripada pencarian kisi atau pencarian serakah. Operator ini menggunakan subproses yang dieksekusi beberapa kali untuk mencari nilai optimal parameter. Operator-parameter Loop juga disediakan untuk mengulangi kombinasi parameter tertentu. Operator pengoptimalan parameter

evolusioner adalah alat yang bermanfaat dalam menemukan nilai optimal untuk parameter yang kompleks dan tidak diketahui dengan jelas.[21].

2.1.7. Optimize Parameter Grid

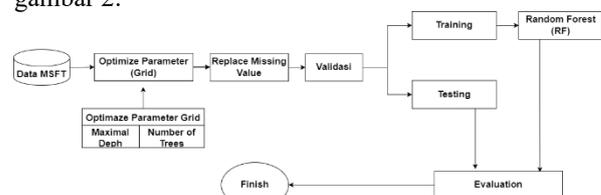
Optimize Parameter Grid adalah metode efektif untuk menentukan nilai parameter optimal dalam subproses operator. Metode ini melibatkan evaluasi performa untuk setiap kombinasi nilai parameter yang dipilih. Dengan mengkomunikasikan nilai parameter optimal melalui port parameter set, metode ini dapat diterapkan pada berbagai jenis operator seperti *Support Vector Classifier*, *Lasso*, *Random Forest*, dan lainnya [22]. *Optimize Parameter Grid* memiliki dua mode: sinkron dan tidak sinkron. Mode sinkron menggunakan daftar kombinasi nilai parameter yang telah ditentukan, sementara mode tidak sinkron menghasilkan semua kombinasi nilai parameter yang mungkin. Dengan menggunakan *Optimize Parameter Grid*, kita dapat secara sistematis dan efisien menemukan nilai parameter terbaik [22].

2.1.8. Evaluasi

Dalam tahap evaluasi ini, nilai RMSE yang mendekati nol terbaik dalam prediksi penutupan harga saham RMSE akan diketahui. Perbandingan hasil RMSE dengan split ratio 0,5 hingga 0,9 dari algoritma *Random Forest* dan algoritma *Random Forest* berbasis *Optimize Parameter Grid* diamati dan *T-test paired two sample for means* menggunakan dilakukan untuk mengetahui adanya perbedaan antara prediksi penutupan harga saham MSFT sebelum optimasi dan setelah optimasi.

2.2. Metode yang Diusulkan

Dalam penelitian ini diusulkan suatu metode untuk prediksi menggunakan *Optimize Parameter Grid* sebagai fitur seleksi parameter dan algoritma *Random Forest* sebagai prediksi penutupan harga saham MSFT. Tahap awal penelitian ini melibatkan pengumpulan dataset MSFT. Dataset kemudian diperiksa dan dilakukan penanganan *missing value*. Data kemudian dipisahkan menjadi *training* dan *testing*. Data *training* digunakan untuk menghasilkan model dengan algoritma *Random Forest*, sedangkan data *testing* digunakan untuk mengukur nilai RMSE. Komparasi algoritma dilakukan untuk menentukan algoritma terbaik dalam pengklasifikasian. Seleksi parameter menggunakan metode *Optimize Parameter Grid*, yang mencari kombinasi parameter terbaik. Fitur-fitur yang dipilih kemudian diprediksi menggunakan algoritma *Random Forest*, yang menggabungkan hasil dari pohon keputusan untuk menghasilkan prediksi akhir. Berikut *proposed method* atau metode yang diusulkan pada gambar 2.



Gambar 2 Proposed Method

3. HASIL DAN PEMBAHASAN

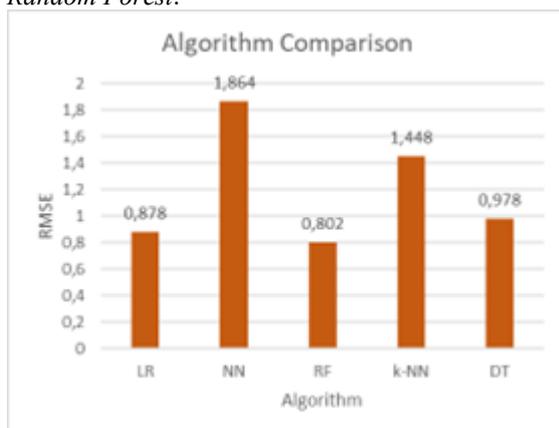
3.1. Pembahasan

Fase ini menyajikan hasil penelitian prediksi harga penutupan saham MSFT. Identifikasi masalah dilakukan untuk menentukan teknik pemodelan yang sesuai. Survei dilakukan untuk prediksi harga penutupan saham MSFT. Dataset publik dari Yahoo Finance digunakan dalam penelitian ini. Preprocessing data melibatkan penyaringan dan pengecekan missing value. Performa 5 algoritma *Linear Regression*, *Decision Tree*, *Neural Network*, *K-Nearest Neighbor*, dan *Random Forest* dibandingkan melalui Validasi data dengan metode *10-Fold Validation*. *Random Forest* (RF) memiliki RMSE terendah (0,802), menunjukkan akurasi prediksi yang tinggi. Data dipisahkan menjadi *training* dan *testing* menggunakan *split validation* dengan variasi rasio. Evaluasi model RF dilakukan untuk harga saham yang belum diketahui sebelumnya dan memberikan informasi tentang tingkat kesalahan prediksi yang optimal. Berikut ini adalah nilai RMSE yang dihasilkan oleh masing-masing algoritma pada tabel 2.

Tabel 2 Komparasi Algoritma

Algoritma	Validation	RMSE
<i>Linear Regression</i>	<i>Cross validation</i>	0,878
<i>Neural Network</i>	<i>Cross validation</i>	1,864
<i>Random Forest</i>	<i>Cross validation</i>	0,802
<i>K-Nearest Neighbor</i>	<i>Cross validation</i>	1,448
<i>Decision Tree</i>	<i>Cross validation</i>	0,978

Sementara itu untuk komparasi algoritma dari semua model ditunjukkan pada grafik di Gambar 3. Pada grafik tersebut dapat dilihat bahwa nilai RMSE terendah mendekati nol ditunjukkan pada algoritma *Random Forest*.



Gambar 3 Grafik RMSE Komparasi Algoritma

Dalam operasi *split validation* menggunakan algoritma *Random Forest*, terdapat perbedaan nilai RMSE tergantung pada rasio pembagian data. Rasio pembagian 0,8 menunjukkan RMSE terendah, menandakan performa terbaik dalam prediksi harga saham menggunakan algoritma *Random Forest*. Tabel 4 menyajikan hasil RMSE algoritma *Random Forest* secara detail. Optimasi parameter memainkan peran penting dalam meningkatkan akurasi prediksi dan

mengurangi nilai RMSE. Melalui perbandingan antara *Optimize Parameter Grid* dan *Optimize Parameter Evolutionary*, dapat ditemukan algoritma yang menghasilkan prediksi yang lebih akurat dengan RMSE yang lebih rendah. Dengan menggunakan algoritma optimasi parameter yang efektif, prediksi saham dapat menjadi lebih baik dan memberikan informasi yang lebih tepat dalam pengambilan keputusan investasi di pasar saham.

Setelah membandingkan *Optimize Parameter Grid* dan *Optimize Parameter Evolutionary* pada optimasi parameter algoritma *Random Forest* (RF), ditemukan hasil menarik. Pada pengujian dengan *split validation*, *Optimize Parameter Grid* pada rasio 0,9 menghasilkan RMSE terendah 0,638, sementara *Optimize Parameter Evolutionary* pada rasio yang sama menghasilkan RMSE terendah 0,651. Komparasi ini menunjukkan bahwa menggunakan metode *Optimize Parameter Grid* dengan parameter yang optimal menghasilkan prediksi yang lebih akurat dibandingkan metode *Optimize Parameter Evolutionary*. Model prediksi dengan *Optimize Parameter Grid* pada rasio 0,9 dapat memberikan estimasi yang lebih akurat dan mendekati nilai sebenarnya dari penutupan harga saham Microsoft Corporation. Untuk melihat hasil yang lebih jelas mengenai optimasi parameter pada Tabel 3:

Tabel 3 Hasil Optimasi Parameter

Valid ation	Rf + Grid			RF + Evolutionary		
	Numb er of Tree	Ma x De ph	RM SE	Num ber of Tree	Ma x De ph	RM SE
<i>Cross</i>	50	10 0	0,76 6	48	23	0,76 6
<i>Split 0,5</i>	100	51	0,78 8	86	56	0,76 9
<i>Split 0,6</i>	70	10 0	0,74 2	94	49	0,71 6
<i>Split 0,7</i>	60	60	0,71 4	80	74	0,72 1
<i>Split 0,8</i>	80	90	0,68 8	66	73	0,74 1
<i>Split 0,9</i>	90	10 0	0,63 8	6	93	0,65 1

3.2. Hasil Penelitian

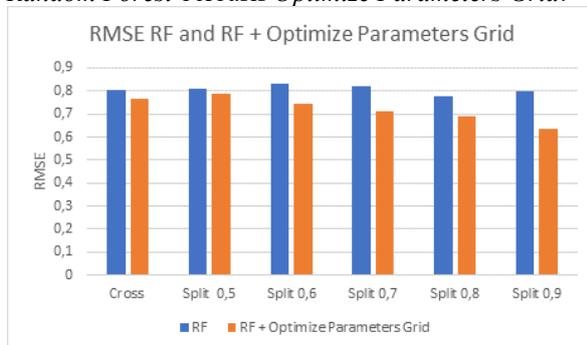
Berdasarkan pengujian yang telah diperoleh pada dataset *microsoft corporation*, dapat diketahui bahwa algoritma *Random Forest* berbasis *Optimize Parameter Grid* efektif dalam memprediksi penutupan harga saham *Microsoft Corporation*. Setelah memilih *Optimize Parameter Grid* sebagai metode optimasi parameter, dilakukanlah komparasi untuk membandingkan nilai RMSE sebelum dan setelah dilakukan optimasi. Hasilnya menunjukkan adanya perubahan yang signifikan dan menuju perbaikan pada nilai RMSE setelah dilakukan optimasi. Hal ini mengindikasikan bahwa penggunaan metode *Optimize*

Parameter Grid telah berhasil meningkatkan akurasi prediksi dalam analisis *Time Series* penutupan harga saham. Perubahan yang menuju ke arah yang lebih baik ini memberikan bukti bahwa optimasi parameter secara efektif dapat membantu mengurangi kesalahan prediksi dan mendekati nilai sebenarnya dari harga saham. Untuk mempermudah dalam memahami perbedaan nilai RMSE antara metode *Random Forest* dan *Random Forest* yang menggunakan *Optimize Parameter Grid* dengan *cross validation* dan *split validation* dari 0,9 sampai 0,5. Tabel 4 dan berikut menunjukkan tabel perbandingan RMSE untuk *Random Forest* dan *Random Forest* berbasis *Optimize Parameter Grid*.

Tabel 4 Perbandingan Algoritma *Random Forest*

No	Rasio	Algoritma	
		RF	RF + <i>Optimize Parameter Grid</i>
1	Cross	0,802	0,766
2	0,5	0,81	0,788
3	0,6	0,832	0,742
4	0,7	0,821	0,714
5	0,8	0,777	0,688
6	0,9	0,798	0,638

Untuk mempermudah dalam memahami perbedaan nilai RMSE antara metode *Random Forest* dan *Random Forest* yang menggunakan *Optimize Parameter Grid* dengan variasi *cross* dan *split validation* dari 0,9 sampai 0,5, dapat membuat grafik. Gambar IV. 3 berikut menunjukkan grafik perbandingan RMSE untuk *Random Forest* dan *Random Forest* berbasis *Optimize Parameters Grid*.



Gambar 4 Grafik Perbandingan RMSE Sebelum dan Sesudah Dioptimasi

T-Test Paired Two Sample menunjukkan perbedaan signifikan ($p < 0,05$) antara nilai RMSE sebelum dan setelah optimasi parameter. Hasil ini mendukung efektivitas metode optimalisasi parameter evolutionary dalam meningkatkan akurasi prediksi harga penutupan saham Microsoft Corporation. Hasil *T-Test Paired Two sample* pada tabel 5.

Tabel 5 Hasil *T-Test Paired Two Sample*

	RMSE sebelum optimasi	RMSE sesudah optimasi
Mean	0,80666667	0,72266667
Variance	0,000367067	0,002993067
Observations	6	6
Pearson Correlation	0,408967734	
Hypothesized Mean Difference	0	
df	5	
t Stat	4,112840222	

P(T<=t) one-tail	0,004619106
t Critical one-tail	2,015048373
P(T<=t) two-tail	0,009238213
t Critical two-tail	2,570581836

3.2.1. Contributions to Knowledge

Penelitian ini menunjukkan inovasi yang luar biasa dengan menggunakan algoritma *Random Forest* untuk analisis *Time Series*, khususnya dalam memprediksi harga penutupan saham Microsoft Corporation. Penelitian ini menggunakan RF untuk menawarkan wawasan baru tentang prediksi dan pemahaman tentang fluktuasi harga saham perusahaan. Analisis aplikasi RF pada data harga saham Microsoft Corporation menghasilkan wawasan baru tentang potensi algoritma ini untuk menganalisis dan memprediksi pergerakan harga saham. Selain itu, penelitian ini menggunakan teknik untuk mengoptimalkan *parameter grid* untuk secara sistematis dan efisien mencari kombinasi parameter optimal untuk algoritma RF. Penggunaan pendekatan ini dalam rangka menganalisis seri harga penutupan Microsoft Corporation menyajikan pendekatan baru untuk memilih dan mengoptimalkan parameter algoritma RF. Kontribusi ini penting dalam menentukan parameter optimal untuk meningkatkan kualitas dan akurasi prediksi algoritma RF dalam analisis *Time Series*.

Secara keseluruhan, penelitian ini menambahkan dua bagian pengetahuan baru. Pertama, penelitian ini meningkatkan pemahaman tentang bagaimana algoritma RF digunakan untuk memprediksi *Time Series* harga saham. Penelitian ini memberikan pemahaman yang lebih mendalam tentang potensi algoritma ini dalam memprediksi pergerakan harga saham melalui penggunaan algoritma RF pada data penutupan harga saham Microsoft Corporation. Kedua, penelitian ini memberikan kontribusi substansial untuk memilih parameter terbaik untuk algoritma RF dalam konteks analisis *Time Series* dengan menggunakan metode *Optimize Parameter Grid*. Efektivitas algoritma RF dalam memprediksi harga saham telah meningkat secara signifikan sebagai hasil dari penelitian ini, yang pada akhirnya dapat meningkatkan preferensi investasi yang lebih baik dan penilaian perdagangan yang lebih akurat.

4. KESIMPULAN

Dalam penelitian ini, analisis *Time Series* harga penutupan saham Microsoft Corporation dilakukan menggunakan algoritma *Random Forest* (RF) dengan *optimize parameter Grid*. Hasil penelitian menunjukkan bahwa penggunaan algoritma RF dalam prediksi harga saham memberikan hasil yang signifikan. Optimisasi parameter menggunakan metode *optimize parameter grid* berhasil meningkatkan kualitas dan akurasi prediksi algoritma RF. Perbedaan signifikan terlihat dalam nilai RMSE, dengan penurunan yang signifikan setelah optimisasi. Analisis statistik dengan *T-Test Paired Two Sample* juga mendukung perbedaan signifikan ini, dengan p-value sebesar 0,009238213. Penelitian ini memberikan bukti

empiris tentang efektivitas algoritma RF dan pentingnya optimisasi parameter dalam meningkatkan kinerja prediksi *Time Series* harga penutupan saham Microsoft Corporation. Hasil ini memiliki implikasi penting dalam pengambilan keputusan investasi dan membuka peluang untuk pengembangan lebih lanjut dalam penggunaan algoritma RF dan optimisasi parameter dalam analisis *Time Series* dan prediksi harga saham.

Saran penelitian yang dapat dieksplorasi adalah pembuatan aplikasi sebagai alat keputusan dalam konteks analisis *Time Series* penutupan harga saham Microsoft Corporation. Dalam penelitian ini, fokus dapat diberikan pada pengembangan aplikasi yang dapat mengintegrasikan model prediksi harga saham berbasis algoritma *Random Forest* dengan data aktual dan indikator ekonomi terkini. Aplikasi ini akan memberikan informasi dan rekomendasi kepada investor untuk pengambilan keputusan investasi yang lebih baik dan akurat. Selain itu, aplikasi tersebut dapat dirancang dengan fitur visualisasi yang menarik dan intuitif, sehingga memudahkan pengguna dalam memahami dan menginterpretasikan hasil prediksi.

5. DAFTAR PUSTAKA

- [1] B. D. Prasetya, F. S. Pamungkas, and I. Kharisudin, "Pemodelan dan Peramalan Data Saham dengan Analisis Time Series menggunakan Python," *PRISMA, Prosiding Seminar Nasional Matematika*, vol. 3, pp. 714–718, 2020, [Online]. Available: <https://journal.unnes.ac.id/sju/index.php/prisma/article/view/38116>
- [2] M. A. D. Suyudi, E. C. Djamil, and A. Maspupah, "Prediksi Harga Saham menggunakan Metode Recurrent Neural Network," *Seminar Nasional Aplikasi Teknologi Informasi (SNATi)*, pp. 1907–5022, 2019.
- [3] H. A. Al Hakim and D. H. Fudholi, "Perbandingan Penggunaan Algoritma Machine Learning pada Prediksi Tren Harga Saham Netflix," *Automata*, vol. 2, no. 2, 2021, [Online]. Available: <https://journal.uin.ac.id/AUTOMATA/article/view/19513>
- [4] I. Networkers, "Microsoft :Pengertian, Sejarah, Dan Perkembangannya," *ID-Networkers*, 2020. <https://www.idn.id/microsoft-pengertian-sejarah-dan-perkembangannya/>
- [5] A. S. Kurniawan, "Prediksi Harga Saham Microsoft Corporation Menggunakan Metode Support Vector Regression.," *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIIK)*, vol. 6, no. 4, pp. 433–440, 2019.
- [6] E. Patriya, "Implementasi Support Vector Machine Pada Prediksi Harga Saham Gabungan (Ihsg)," *Jurnal Ilmiah Teknologi dan Rekayasa*, vol. 25, no. 1, pp. 24–38, 2020, doi: 10.35760/tr.2020.v25i1.2571.
- [7] Y. Ramdhani and A. Mubarak, "Analisis Time Series Prediksi Penutupan Harga Saham Antm.Jk Dengan Algoritma SVM Model Regresi," *Jurnal Responsif: Riset Sains dan Informatika*, vol. 1, no. 1, pp. 77–82, 2019.
- [8] M. E. Bastian, B. Rahayudi, and D. E. Ratnawati, "Prediksi Trend Harga Saham Jangka Pendek berdasarkan Fitur Technical Analysis dengan menggunakan Algoritma Random Forest," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 5, no. 10, pp. 4536–4542, 2021, [Online]. Available: <http://j-ptiik.ub.ac.id>
- [9] D. A. Rezaldi and Sugiman, "Peramalan Metode ARIMA Data Saham PT . Telekomunikasi Indonesia," *PRISMA, Prosiding Seminar Nasional Matematika*, vol. 4, pp. 611–620, 2021.
- [10] A. A. Parikesit, R. Nurdiansyah, and D. Agustriawan, "Penerapan Pendekatan Machine Learning Pada Pengembangan Basis Data Herbal Sebagai Sumber Informasi Kandidat Obat Kanker," *Jurnal Teknologi Industri Pertanian*, vol. 29, no. 2, pp. 175–182, 2019, doi: 10.24961/j.tek.ind.pert.2019.29.2.175.
- [11] Mr. T. Phase and Dr. Prof. S. Patil, "Predict the Level of Income using Random Forest Classifier," *Int J Res Appl Sci Eng Technol*, vol. 7, no. 12, pp. 558–561, 2019, doi: 10.22214/ijraset.2019.12090.
- [12] A. Pratomo, R. F. Umbara, and A. A. Rohmawati, "Prediksi Pergerakan Harga Saham Dengan Metode Random Forest Menggunakan Trend Deterministic Data Preparation (Studi Kasus Saham Perusahaan Pt Astra International Tbk, Pt Garuda Indonesia Tbk, Dan Pt Indosat Tbk)," *eProceedings of Engineering*, vol. 6, no. 1, pp. 2545–2556, 2019.
- [13] A. Hermawan, I. W. Mangku, N. K. K. Ardana, and H. Sumarno, "Analisis Support Vector Regression Dengan Algoritma Grid Search Untuk Memprediksi Harga Saham," *MILANG Journal of Mathematics and Its Applications*, vol. 18, no. 1, pp. 41–60, 2022, doi: 10.29244/milang.18.1.41-60.
- [14] N. Manullang, R. W. Sembiring, I. Gunawan, I. Parlina, and Irawan, "Implementasi Teknik Data Mining untuk Prediksi Peminatan Jurusan Siswa Menggunakan Algoritma C4.5," *Jurnal Ilmu Komputer dan Teknologi*, vol. 2, no. 2, pp. 1–5, 2021, doi: 10.35960/ikomti.v2i2.700.
- [15] D. Muriyatmoko, "Pengaruh indeksasi doaj terhadap sitasi pada jurnal terakreditasi sinta menggunakan analisis regresi linier," *Jurnal SimanteC*, vol. 7, no. 1, pp. 31–38, 2018.
- [16] J. Han and M. Kamber, *Data Mining: Concepts and Techniques, Second Edition (The Morgan Kaufmann Series in Data Management*

- Systems*), 2nd ed., vol. 340. Morgan kaufmann, 2006.
- [17] Y. Ramdhani, A. Mubarok, S. Hidayatullah, and W. Wiguna, "Attribute Optimization: Genetic Algorithms and Neural Network for Voice Analysis Classification of Parkinson's Disease," *SCITEPRESS – Science and Technology Publications*, no. Icri 2018, pp. 3074–3079, 2020, doi: 10.5220/0009947030743079.
- [18] N. T. Romadloni, I. Santoso, and S. Budilaksono, "Perbandingan Metode Naive Bayes, Knn Dan Decision Tree Terhadap Analisis Sentimen Transportasi Krl Commuter Line," *Jurnal IKRA-ITH Informatika: Jurnal Komputer dan Informatika*, vol. 3, no. 2, pp. 1–9, 2019.
- [19] C. Cahyaningtyas, Y. Nataliani, and I. R. Widiasari, "Analisis Sentimen Pada Rating Aplikasi Shopee Menggunakan Metode Decision Tree Berbasis SMOTE," *AITI*, vol. 18, no. 2, pp. 173–184, 2021, doi: 10.24246/aiti.v18i2.173-184.
- [20] Yahya and W. P. Hidayanti, "Penerapan Algoritma K-Nearest Neighbor Untuk Klasifikasi Efektivitas Penjualan Vape (Rokok Elektrik) pada 'Lombok Vape On,'" *Infotek : Jurnal Informatika dan Teknologi*, vol. 3, no. 2, pp. 104–114, 2020, doi: 10.29408/jit.v3i2.2279.
- [21] Rapidminer, "Optimize Parameters (Evolutionary)," *RapidMiner Studio Core*. https://docs.rapidminer.com/9.9/studio/operators/modeling/optimization/parameters/optimize_parameters_evolutionary.html (accessed Jul. 11, 2023).
- [22] RapidMiner, "Optimize Parameters (Grid)," *RapidMiner Documentation*, 2022. https://docs.rapidminer.com/9.10/studio/operators/modeling/optimization/parameters/optimize_parameters_grid.html (accessed Jul. 19, 2023).