EVALUASI KINERJA MODEL PREDIKSI PERGERAKAN HARGA SAHAM INDF.JK MENGGUNAKAN METODE *RANDOM FOREST* DENGAN OPTIMASI *GRID SEARCH CROSS VALIDATION*

DELLA ZARIA
H1091211002

SKRIPSI

PROGRAM STUDI STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS TANJUNGPURA
PONTIANAK
2025

EVALUASI KINERJA MODEL PREDIKSI PERGERAKAN HARGA SAHAM INDF.JK MENGGUNAKAN METODE *RANDOM FOREST* DENGAN OPTIMASI *GRID SEARCH CROSS VALIDATION*

DELLA ZARIA
H1091211002

Skripsi
Sebagai salah satu syarat untuk memperoleh gelar
Sarjana Statistika pada Program Studi Statistika

PROGRAM STUDI STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS TANJUNGPURA
PONTIANAK
2025

LEMBAR PENGESAHAN

Judul Tugas Akhir	: Evaluasi Kinerja Model Prediksi Pergerakan Harga Saham INDF.JK Menggunakan Metode *Random Forest* Dengan Optimasi *Grid Search Cross Validation*
Nama Mahasiswa	: Della Zaria
NIM	: H1091211002
Jurusan/Program Studi	: Matematika/Statistika
Tanggal Lulus	: 21 Januari 2025
SK Pembimbing	: No. 2582/UN22.8/TD.06/2024/Tanggal 2 September 2024
SK Penguji	: No. 5555/UN22.8/TD.06/2024/Tanggal 31 Desember 2024

Dosen Pembimbing

Pembimbing I	Pembimbing II
Dr. Evy Sulistianingsih, M.Sc.	Shantika Martha, M.Si.
NIP. 198502172008122006	NIP. 198403082008122003

Dosen Penguji

Ketua Penguji	Anggota Penguji
Neva Satyahadewi, M.Sc.,CRA.,CRP.,CRMP.	Nurfitri Imro'ah, M.Si.
NIP. 198212042005012001	NIP. 198907182019032021

Pimpinan Sidang (merangkap anggota penguji)	Sekretaris Sidang (merangkap anggota penguji)
Dr. Evy Sulistianingsih, M.Sc.	Shantika Martha, M.Si.
NIP. 198502172008122006	NIP. 198403082008122003

Mengesahkan
Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam
Universitas Tanjungpura

Prof. Dr. Gusrizal, M.Si.
NIP. 197108022000031001

PERNYATAAN INTEGRITAS AKADEMIK

Saya yang bertanda tangan di bawah ini:

Nama : Della Zaria

NIM : H1091211002

Program Studi/ Jurusan : Matematika/Statistika

Fakultas : Matematika dan Ilmu Pengetahuan Alam

dengan ini menyatakan bahwa dokumen ilmiah Tugas Akhir yang disajikan ini tidak mengandung unsur pelanggaran integritas akademik sesuai Peraturan Menteri Pendidikan, Kebudayaan, Riset, Dan Teknologi Republik Indonesia Nomor 39 Tahun 2021. Apabila di kemudian hari dokumen ilmiah Tugas Akhir ini mengandung unsur pelanggaran integritas akademik sesuai ketentuan perundangan tersebut, maka saya bersedia menerima sanksi akademik dan/atau sanksi hukum yang berlaku.

Demikian pernyatan ini untuk digunakan sebagaimana mestinya.

Pontianak, 21 Januari 2025



Della Zaria
H1091211002

Evaluasi Kinerja Model Prediksi Pergerakan Harga Saham INDF.JK Menggunakan Metode *Random Forest* Dengan Optimasi *Grid Search Cross Validation*

Abstrak

Investasi pada instrumen keuangan telah menunjukkan pertumbuhan yang signifikan seiring berjalannya waktu, termasuk saham yang seringkali menjadi pilihan utama para investor untuk menanamkan uang. Namun sayangnya, perubahan harga saham yang fluktuatif dan kerugian yang dapat ditimbulkan menjadi permasalahan khusus bagi calon investor untuk menentukan waktu yang tepat. Penelitian ini bertujuan untuk mengevaluasi kinerja model prediksi pergerakan harga saham PT Indofood Sukses Makmur Tbk (INDF.JK) pada 30 hari mendatang menggunakan metode Random Forest dan optimasi *Grid Search Cross Validation* (CV) guna membantu proses pengambilan keputusan berdasarkan hasil kinerja. Data yang digunakan adalah harga penutupan saham INDF.JK periode 02 Januari 2014 sampai 29 Desember 2023 yang diolah menjadi delapan jenis indikator teknikal saham. Tahapan yang dilalui penelitian ini diantaranya statistik deskriptif, *preprocessing,* penentuan variabel fitur dan target, pembagian data, pembentukan model tanpa dan dengan optimasi, pengujian disertai evaluasi kinerja, dan perbandingan model yang terbentuk. Hasil menunjukkan model prediksi dari *Random Forest* dengan optimasi *Grid Search CV* memiliki performa yang sangat baik dengan keakuratan sebesar 90,8%. Nilai *precision*, *sensitivity*, dan *specificity* model ini berturut-turut sebesar 90,8%, 91,5%, dan 90,1%. Penggunaan optimasi *Grid Search CV* memberikan hasil kinerja model yang lebih baik dibanding tanpa optimasi untuk memprediksi pergerakan harga penutupan saham INDF.JK yang ditunjukkan oleh kenaikan indikator kinerja. *Relative Strength Index* (RSI) menjadi variabel dengan performa yang paling baik bagi model prediksi sehingga dapat dijadikan pertimbangan utama pengambilan keputusan proses jual beli saham INDF.JK.

Kata kunci: Saham, *Random Forest*, *Grid Search, Cross Validation*

Performance Evaluation of The INDF.JK Stock Price Movement Prediction Model Using Random Forest Method with Grid Search Cross Validation Optimization

Abstract

Investment in financial instruments in Indonesia has shown significant growth over time, with stocks often being the first choice for investors to invest money. Unfortunately, fluctuations in stock prices and losses caused by wrong decisions are special problems for investors when determining the right time. This study aims to evaluate the performance of the prediction model for PT Indofood Sukses Makmur Tbk (INDF.JK) stock price movement in the next 30 days using the Random Forest method and Grid Search Cross Validation (CV) optimization to help the decision-making process based on performance results. The data used is the INDF.JK stock's closing price in January 2, 2014 to December 29, 2023, which is processed into eight types of stock technical indicators. The research pipeline includes descriptive statistics, preprocessing, feature and target variables determination, data split, model formation without and with optimization, testing accompanied by performance evaluation, and comparison of the formed model. The results show that the prediction model from Random Forest with Grid Search CV optimization has excellent performance with an accuracy of 90.8%. The precision, sensitivity, and specificity values of this model are 90.8%, 91.5%, and 90.1%, respectively. The use of Grid Search CV optimization provides better model performance results than without optimization to predict the closing price movement of INDF.JK stock indicated by the increase of performance indicators. Relative Strength Index (RSI) is the variable with the best performance for the model so that can be primary consideration in making decisions on buy and sell process of INDF.JK stock.

Keywords: *Stock, Random Forest, Grid Search, Cross Validation*

PRAKATA

Puji dan syukur atas kehadirat Allah SWT atas limpahan rahmat dan karunianya sehingga penulis dapat menyelesaikan tugas akhir yang berjudul "**Evaluasi Kinerja Model Prediksi Pergerakan Harga Saham INDF.JK Menggunakan Metode *Random Forest* Dengan Optimasi *Grid Search Cross Validation***". Sholawat serta salam tidak lupa senantiasa dijunjungkan kepada Nabi Muhammad SAW.

Penulisan tugas akhir ini bertujuan memenuhi syarat dalam memperoleh gelar Sarjana Statistika pada Prodi Statistika Jurusan Matematika Fakultas MIPA Universitas Tanjungpura dan sebagai apresiasi penulis terhadap diri sendiri atas perjalanan panjang yang telah ditempuh. Penyusunan tugas akhir ini tentu tidak terlepas dari dukungan berbagai pihak. Ucapan terima kasih penulis sampaikan kepada:

1. Allah SWT, yang telah memberikan kekuatan dan kelancaran atas setiap langkah yang sudah, sedang, dan akan dihadapi, yang telah menyisipkan limpahan kebahagiaan dan kebaikan dalam hidup penulis, sedari awal hingga akhir.
2. Bapak Zailani selaku ayah dan Ibu Satriani selaku mama, kedua orang tua tercinta penulis yang luar biasa hebatnya, yang selalu percaya atas setiap langkah yang penulis ambil, yang telah memberikan ribuan doa, semangat, dukungan, dan menjadi motivasi terbesar bagi penulis untuk menyelesaikan tugas akhir ini.
3. Adik penulis, Mitha Zaria, yang telah menemani hari-hari penulis di kamar dan menjadikan kamar sebagai tempat ternyaman menyusun tugas akhir bagi penulis.
4. Ibu Dr. Evy Sulistyaningsih, M.Sc. selaku Dosen Pembimbing Pertama yang telah membimbing penulis dengan sepenuh jiwa, memberi arahan, saran, dan masukan, serta semangat kepada penulis untuk menyelesaikan tugas akhir ini dengan baik.
5. Ibu Shantika Martha, M.Si. selaku Dosen Pembimbing Kedua yang telah membimbing penulis dengan sepenuh hati, memberi arahan, saran, dan masukan, serta semangat kepada penulis untuk menyelesaikan tugas akhir ini dengan baik.

6. Ibu Wirda Andani, M.Si. selaku Dosen Prodi Statistika yang telah bersedia turut serta membersamai penulis dalam menyelesaikan tugas akhir ini, memberikan arahan, saran, dan masukan terhadap implementasi metode dan tugas akhir penulis.
7. Neva Satyahadewi, M.Sc.,CRA.,CRP.,CRMP. selaku Dosen Penguji Pertama yang telah memberikan beragam arahan, saran, dan masukan untuk tugas akhir ini.
8. Nurfitri Imro'ah, M.Si. selaku Dosen Penguji Kedua yang telah memberikan beragam arahan, saran, dan masukan untuk tugas akhir ini.
9. Ancugh Family selaku teman seperjuangan sedari maba, diantaranya Lusi, Muti, Putri, Hera, Ghina, Siska, Wahida, Nanda, dan Ulya, yang selalu menjadi rumah ternyaman penulis untuk tertawa, yang senantiasa menjadi warna-warni perkuliahan penulis hingga penyelesaian tugas akhir ini.
10. Teman-teman GenBI Kalbar, yang turut menambahkan warna dalam kehidupan penulis dengan limpahan kebahagiaan, kebermanfaatan yang ditebar, dukungan emosional, dan ribuan cerita serunya yang selalu mewarnai isi kepala penulis.
11. Teman-teman semasa sekolah penulis yang masih turut membersamai (Lipad Smansa beserta grup-grup lainnya yang selalu menyenangkan ketika liburan tiba).
12. Della Zaria selaku penulis, yang telah mengusahakan tugas akhir ini dengan sekuat tenaga, yang selalu percaya bahwa tugas akhir ini pasti bisa diselesaikan, dan yang senantiasa memaknai setiap *progress* yang dilakukan, meski kecil sekalipun.
13. Rekan Expos'21 dan berbagai pihak yang telah turut membantu dan membersamai penulis selama proses penyusunan skripsi dan belum dapat disebutkan satu-persatu.

Penulis berharap tugas akhir ini mampu membawa kebermanfaatan, tidak hanya bagi penulis, tetapi juga bagi sekitar. Penulis sangat menerima kritik dan saran yang membangun demi mencapai kebermanfaatan maksimal dari penulisan tugas akhir ini.

Pontianak, 21 Januari 2025



Della Zaria

DAFTAR ISI

PRAKATA ..... **vi**
DAFTAR GAMBAR ..... **x**
DAFTAR TABEL ..... **xi**
DAFTAR SIMBOL ..... **xii**
BAB I PENDAHULUAN **1**
1.1 Latar Belakang 1
1.2 Rumusan Masalah 4
1.3 Tujuan Penelitian 4
1.4 Batasan Masalah 5
1.5 Tinjauan Pustaka 6
1.6 Metodologi Penelitian 8
BAB II LANDASAN TEORI **10**
2.1 Saham 10
2.2 *Data Mining* 13
2.3 Prediksi 15
2.4 Evaluasi Kinerja Model 16
BAB III *RANDOM FOREST – GRID SEARCH CROSS VALIDATION (CV)* **18**
3.1 *Random Forest* 18
3.2 *Grid Search Cross-Validation (CV)* 21
3.2.1 *Grid Search* 22
3.2.2 *K-Fold Cross Validation* 23
BAB IV HASIL DAN PEMBAHASAN **26**
4.1 Deskripsi Data 26
4.2 *Preprocessing* Data 27
4.3 Penentuan Variabel Fitur dan Target 30
4.3.1 Variabel Fitur 30
4.3.2 Variabel Target 31

4.4 Pembagian Data 34
4.5 Pembentukan dan Evaluasi Kinerja Model dari Metode *Random Forest* tanpa Optimasi *Grid Search CV* 36
4.6 Pembentukan dan Evaluasi Kinerja Model dari Metode *Random Forest* dengan Optimasi *Grid Search CV* 40
4.7 Perbandingan Hasil Evaluasi Kinerja Model Prediksi 45
BAB V PENUTUP 47
5.1 Kesimpulan 47
5.2 Saran 47
DAFTAR PUSTAKA 49

DAFTAR GAMBAR

Gambar 1. 1 *Flowchart* Penelitian 9
Gambar 3.1 Skema Kerja *Random Forest* 19
Gambar 3. 2 Skema Kerja *Grid Search* 23
Gambar 3. 3 Skema Kerja *K-Fold Cross Validation* 24
Gambar 4. 1 Grafik Pergerakan Harga Penutupan Saham INDF.JK pada Tahun 2014-2023 27
Gambar 4. 2 Perbandingan Pergerakan Harga Saham INDF.JK Tanpa dan Dengan Pemulusan 29
Gambar 4. 3 Pohon Pertama dari Model yang Menggunakan Metode *Random Forest* Tanpa *Grid Search CV* 36
Gambar 4. 4 Pohon Pertama dari Model yang Menggunakan Metode *Random Forest* Dengan *Grid Search CV* 42
Gambar 4. 5 Tingkat Kepentingan Variabel Model dari Metode *Random Forest* Tanpa dan Dengan Optimasi *Grid Search CV* 47

DAFTAR TABEL

Tabel 2. 1 *Confusion Matrix* 16
Tabel 4. 1 Statistik Deskriptif Data Harga Penutupan Saham INDF.JK dari 02 Januari 2014 sampai 29 Desember 2023 26
Tabel 4. 2 Hasil Pemulusan Eksponensial Harga Saham INDF.JK 28
Tabel 4. 3 Daftar Variabel Fitur 30
Tabel 4. 4 Variabel Fitur dari Hasil Perhitungan Indikator Teknikal Saham 30
Tabel 4. 5 Data Variabel Fitur Hasil Pembersihan 31
Tabel 4. 6 Variabel Target Hasil Perhitungan Harga Penutupan Saham INDF.JK 32
Tabel 4. 7 Hasil Data Variabel Target Setelah *Data Cleansing* 33
Tabel 4. 8 Jumlah Data Tiap Kelas Variabel Target 34
Tabel 4. 9 Jumlah Pembagian Data Per Kelas 35
Tabel 4.10 *Confusion Matrix* Hasil Prediksi Model Oleh Metode *Random Forest* Tanpa *Grid Search CV* 37
Tabel 4. 11 Tingkat Kepentingan Variabel Relatif Model Tanpa *Grid Search CV* 39
Tabel 4. 12 Hasil *Hyperparameter Tuning Grid Search CV* 40
Tabel 4. 13 *Confusion Matrix* Hasil Prediksi Model Oleh Metode *Random Forest* Dengan *Grid Search CV* 43
Tabel 4. 14 Tingkat Kepentingan Variabel Relatif Model Dengan *Grid Search CV* 44
Tabel 4. 15 Perbandingan Nilai Indikator Kinerja Model dari Metode *Random Forest* Tanpa dan Dengan Optimasi *Grid Search CV* 45

DAFTAR SIMBOL

SMA_i : Rata-rata pergerakan harga saham pada hari ke-i, $i = 1,2, \dots, N$

C_i : Harga penutupan saham pada hari ke- i

C_{i-1} : Harga penutupan saham pada 1 hari sebelumnya

C_{i-2} : Harga penutupan saham pada 2 hari sebelumnya

$C_{i-(n-1)}$: Harga penutupan saham pada $(n-1)$ hari sebelumnya, $n = 1,2, \dots, N$

n : Jumlah periode waktu

N : Jumlah data

EMA_{i-1} : Nilai EMA pada 1 hari sebelumnya

EMA_n : Nilai EMA selama n hari

u_n : Kenaikan terakhir dalam periode waktu n

d_n : Penurunan terakhir dalam periode waktu n

k : Nilai dari jumlah pohon keputusan

m : Jumlah kelas target

p_b : Probabilitas kemunculan kelas ke-b pada partisi D, $b = 1,2, \dots, m$

v : Jumlah partisi

D_j : Total partisi ke- j, $j = 1,2, \dots, v$

D : Jumlah baris pada seluruh partisi

$Gini\ (h)^k$: Indeks gini untuk variabel independen X_h pada pohon ke-*k, h = 1, 2, ..., p*

p : Jumlah variabel independen

F_i : Nilai ramalan pada hari ke-i

α : Parameter pemulusan $(0 < \alpha < 1)$

X_i : Nilai aktual pada hari ke-i

F_{i-1} : Nilai ramalan pada 1 hari sebelumnya

$Target_i$: Kelas variabel target dari data hari ke-i

Sign : Penggunaan tanda kelas, yakni 1 atau -1

C_{i+30} : Harga penutupan saham untuk 30 hari ke depan

BAB I
PENDAHULUAN

1.1 Latar Belakang

Seiring perkembangan waktu, investasi pada instrumen keuangan di Indonesia memperlihatkan pertumbuhan yang memuaskan secara signifikan. Bursa Efek Indonesia (BEI) melaporkan bahwa jumlah investor pasar modal yang mencakup investor saham, obligasi, dan reksa dana per tahun 2023, berhasil menyentuh angka 12,16 juta investor dengan peningkatan sebanyak 1,85 juta investor dari tahun sebelumnya (idx.co.id, 2023). Terlebih lagi, perkembangan teknologi dan informasi dari hari ke hari tentu kian membawa berbagai kemudahan aksesibilitas bagi setiap orang untuk berinvestasi. Perkembangan ini menggerakkan perubahan model investasi pada instrumen keuangan dari tradisional menjadi digital sehingga calon investor tidak hanya mudah dalam mengakses, tetapi juga mampu mengamati pergerakan harga secara *real time* dan memperoleh keputusan untuk berinvestasi dengan lebih singkat (Rais, Khairi, dan Hidayat, 2023).

Beragam jenis investasi pada instrumen keuangan yang dapat diakses di pasar modal, salah satunya adalah saham. Saham merupakan alat investasi yang diyakini dapat memperoleh hasil yang menguntungkan sehingga seringkali menjadi pilihan utama investor dalam menanamkan uang (Syafira & Saepudin, 2023). Investasi saham diunggulkan oleh kapabilitasnya dalam menghasilkan keuntungan (*rate of return*) kepada investor dengan nilai yang tidak terhingga (Kurniawan, 2019). Tak heran, jumlah investor saham di Indonesia tercatat telah melampaui angka 5,7 juta *Single Investor Identification* (SID) hingga pertengahan tahun 2024 (idx.co.id, 2024). Besarnya angka ini tentunya masih diiringi peningkatan jumlah seiring waktu, terutama bagi calon investor saham yang berminat untuk turut berinvestasi namun belum merealisasikannya, yakni melakukan pembelian saham di pasar modal.

Namun sayangnya, keputusan untuk melakukan pembelian saham pada satu waktu pilihan bukanlah hal yang mudah. Terdapat permasalahan yang seringkali dihadapi oleh calon investor untuk melakukan pembelian saham, yakni pergerakan harga saham yang mengalami perubahan secara fluktuatif dalam kurun waktu singkat (Pratomo, Umbara, dan Rohmawati, 2019). Perubahan pergerakan harga saham dapat terjadi akibat berbagai faktor, di antaranya situasi perusahaan, situasi perekonomian negara bahkan dunia secara menyeluruh, sampai sentimen pasar (Valiant dkk., 2019). Meskipun begitu, kondisi pergerakan saham inilah yang seharusnya dijadikan sebagai salah satu pertimbangan dengan cukup matang sebelum memutuskan berinvestasi pada suatu saham (Setiawan & Mulyani, 2020). Jika kesalahan pengambilan keputusan dilakukan, maka secara otomatis akan mengakibatkan kerugian bagi pemilik saham. Kesalahan yang tidak diharapkan ini semestinya dapat diminimalisir oleh penganalisaan harga saham dengan melakukan prediksi pergerakan harga saham pada waktu mendatang berdasarkan nilai-nilai di waktu sekarang. Analisis prediksi pergerakan harga saham dapat dilakukan dengan menggunakan data historis harian saham (*open, close, high, low,* dan *volume*) dan data indikator teknikal (*moving average, Relative Strength Index* (RSI), dan lain-lain) (Priyatno dkk., 2023). Dengan adanya analisis ini, calon investor mampu menghasilkan keputusan pilihan waktu yang tepat untuk berinvestasi sehingga dapat memaksimalkan keuntungan dan meminimalkan kerugian seperti yang diharapkan.

Salah satu metode pembelajaran yang mendukung kemudahan analisis prediksi pergerakan harga saham adalah *Data Mining*. *Data Mining* merupakan ilmu pembelajaran yang dimanfaatkan dalam memberikan gambaran dan mengungkap informasi baru dari sekumpulan data (Yulianti, Soesanto, dan Sukmawaty, 2022). *Data Mining* mampu menjalankan prediksi yang akurat pada berbagai kasus (Aeni, 2020).

Salah satu algoritma bagian dari *Data Mining* yang seringkali digunakan untuk menganalisis data dengan tujuan prediksi dan klasifikasi adalah *Random Forest*. Algoritma ini mampu menyelesaikan berbagai permasalahan terkait klasifikasi data, seperti klasifikasi kelayakan kredit dan klasifikasi penyakit daun tomat, berlaku juga

pada berbagai kasus prediksi, seperti prediksi harga bitcoin, harga ponsel, dan curah hujan (Ismanto & Novalia, 2021). *Random Forest* mempunyai beberapa kelebihan dalam sistem kerjanya, yakni mampu menghasilkan luaran klasifikasi dan prediksi yang baik dengan besar kemungkinan kesalahan yang minim, mampu menangani data *training* dalam jumlah yang besar dengan efektif, dan diyakini efisien dalam mengendalikan permasalahan *missing value* pada data yang dianalisis (Lestari & Sirodj, 2022). Sebagai pengembangan dari *Decision Tree*, algoritma ini efisien dalam menganalisis data dan memberikan hasil terbaik dengan memperbaiki kinerja metode-metode pada *Decision Tree* yang rentan dengan gejala *overfitting* (Amini dkk., 2022). Kriteria data yang dibutuhkan untuk menghasilkan kinerja optimal dari metode ini adalah jumlah kelas pada variabel target sebaiknya seimbang, data bersih dari *missing value,* dan data inputan harus berupa variabel numerik sebagai kriteria khususnya.

Namun, dibalik semua keunggulannya, *Random Forest* memerlukan variasi *hyperparameter* yang terdiri atas jumlah estimator, *depth tree*, kriteria pemilihan fitur, dan parameter lainnya yang harus diujikan ke dalam model untuk memperoleh penggunaan paramater dengan hasil paling optimal (Azhar & Pardede, 2021). Adapun metode yang paling efisien dalam mencari kombinasi parameter terbaik adalah menyesuaikan *hyperparameter* dengan pengoptimalan menggunakan *hyperparameter tuning* (George & Sumathi, 2020), salah satunya yaitu *Grid Search Cross Validation (CV)*. Optimasi ini mampu menemukan kombinasi *hyperparameter* terbaik, di mana setiap kombinasi parameter yang ada pada *grid* akan divalidasi dengan *Cross Validation* dan selanjutnya dibangkitkan dan dievaluasi oleh *Grid Search CV* (Aina, Nastiti, dan Aditya, 2024). Proses validasi oleh *Cross Validation* ini akan memastikan model yang terbentuk oleh pilihan *hyperparameter* dari *Grid Search* mewakili keseluruhan pola dalam dataset dan tidak bergantung pada pembagian data (Sugiman & Purnomo, 2024).

Pada penelitian ini, *Random Forest* dengan optimasi parameter *Grid Search CV* diaplikasikan pada PT Indofood Sukses Makmur Tbk (INDF.JK) sebagai sasaran studi kasus analisis prediksi pergerakan harga saham. PT Indofood Sukses Makmur Tbk

merupakan industri terkemuka di Indonesia yang bergerak pada sektor barang konsumen primer dengan bentuk *Total Food Solution* (Utami & Hardana, 2022). Sektor industri ini memiliki peluang yang cukup besar untuk senantiasa diandalkan masyarakat dalam kehidupan sehari-hari sehingga prospeknya dapat dikatakan baik tidak hanya di masa kini, tetapi juga di masa depan (Dewi & Siska, 2024). Dengan demikian, evaluasi kinerja model prediksi pergerakan harga saham INDF.JK ini diharapkan mampu menilai keoptimalan model prediksi yang dibentuk dalam memberikan gambaran terkait pergerakan harga saham INDF.JK sehingga dapat membantu proses pengambilan keputusan jual beli saham INDF.JK.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah dipaparkan, maka rumusan masalah yang hendak dijawab melalui penelitian ini adalah sebagai berikut.

1. Bagaimana evaluasi kinerja model prediksi pergerakan harga saham INDF.JK dari pengaplikasian metode *Random Forest* tanpa optimasi *Grid Search CV*?
2. Bagaimana evaluasi kinerja model prediksi pergerakan harga saham INDF.JK dari pengaplikasian metode *Random Forest* dengan optimasi *Grid Search CV*?
3. Bagaimana perbandingan evaluasi kinerja kedua model dalam memprediksi pergerakan harga saham INDF.JK?

1.3 Tujuan Penelitian

Berdasarkan rumusan masalah yang dirumuskan, tujuan yang hendak dicapai pada penelitian ini diantaranya:

1. Mengevaluasi kinerja model prediksi pergerakan harga saham INDF.JK dari pengaplikasian metode *Random Forest* tanpa optimasi *Grid Search CV*.
2. Mengevaluasi kinerja model prediksi pergerakan harga saham INDF.JK dari pengaplikasian metode *Random Forest* dengan optimasi *Grid Search CV*.
3. Membandingkan evaluasi kinerja kedua model dalam memprediksi pergerakan harga saham INDF.JK.

1.4 Batasan Masalah

Untuk menyusun tulisan yang lebih terarah, dilakukan pembatasan masalah pada penelitian ini diantaranya sebagai berikut.

1. Data yang digunakan dalam penelitian ini merupakan data sekunder berupa data historis harga saham harian PT Indofood Sukses Makmur Tbk (INDF.JK) periode 02 Januari 2014 sampai 29 Desember 2023 di mana total data sebanyak 2.482 yang didapat dari website www.yahoofinace.com.
2. Variabel independen yang digunakan dalam penelitian ini merupakan indikator teknikal saham, yang terdiri dari *Simple Moving Average* (SMA), *Exponential Moving Average* (EMA), *Moving Average Convergence/Divergence* (MACD), dan *Relative Strength Index* (RSI) di mana SMA_5 (X1), SMA_10 (X2), SMA_15 (X3), SMA_30 (X4), EMA_9 (X5), MACD (X6), MACD_Signal (X7), dan RSI (X8). Adapun variabel dependen (Y) yang ditetapkan adalah harga penutupan saham yang telah dikategorikan dalam dua kelas, yakni kelas naik (+1) dan turun (-1).
3. Pemulusan eksponensial yang dilakukan pada data harga penutupan saham menggunakan metode *Single Exponential Smoothing (SES).*
4. Analisis prediksi pergerakan harga saham yang dilakukan pada penelitian ini ditujukan untuk memprediksi pergerakan harga saham dalam jangka pendek, yakni dalam waktu 30 hari mendatang.
5. Pembagian data *training* dan data *testing* dalam penelitian ini menggunakan komposisi 80:20.
6. Parameter *Random Forest* yang digunakan pada penelitian ini diantaranya *mtry* (jumlah variabel yang dipilih pada setiap split) senilai 2, 3, 4, 8 dan *ntree* (jumlah pohon dalam *forest*) senilai 50, 100, 200, 500, 1000.
7. Penelitian ini menggunakan *10-Fold Cross Validation* dalam memvalidasi model yang terbentuk.

1.5 Tinjauan Pustaka

Baihaqi dkk. (2023) melakukan penelitian terkait pengimplementasian algoritma *Random Forest* untuk memprediksi harga saham didasarkan pada data historis dengan menggunakan optimasi *Grid Search CV*. Data yang digunakan untuk melatih dan menguji model adalah data saham PT Telkom Indonesia periode Januari tahun 2014 hingga Oktober tahun 2023 yang sudah di*preprocessing* sebelumnya. Pembangunan model prediksi dijalankan dengan menggunakan metode *Random Forest* di mana *Grid Search CV* dengan *5-Fold Cross Validation* selanjutnya difungsikan dalam memilih kombinasi *hyperparameter* terbaik pada model yang hendak dibentuk. Pengukuran evaluasi kinerja model menggunakan nilai *Mean Squared Error* (MSE), *Root Mean Squared Error* (RMSE), dan *Mean Absolute Error* (MAE) yang dihitung untuk mengetahui keakuratan model memprediksi data. Didapatkan hasil yang terhitung cukup akurat dalam memprediksi data, berturut-turut senilai 0,984, 0,992, dan 0,277.

Pengoptimalan kinerja metode *Random Forest* dengan *hyperparameter tuning Grid Search CV* juga dilakukan oleh penelitian Sitanggang dan Sitompul (2024) yang mengaplikasikan metode tersebut untuk melakukan deteksi awal kelangsungan hidup pasien gagal jantung. Penelitian ini menyertakan *hyperparameter tuning Random Search* sebagai bahan komparasi *Grid Search CV* dalam mengoptimalkan kinerja model. Hasil akurasi, *recall*, dan spesifisitas dari metode *Random Forest* tanpa *hyperparameter tuning* apapun, *Random Forest* dengan *Random Search CV*, dan *Random Forest* dengan *Grid Search CV* dibandingkan satu sama lain. Didapatkan hasil kombinasi *Random Forest* dengan *Grid Search CV* memiliki kinerja paling optimal dengan tingkat akurasi tertinggi sebesar 85%, *recall* 57,89%, dan spesifisitas 97,56%. Pemanfaatan *hyperparameter tuning* baik *Random Search CV*, maupun *Grid Search CV* pada penelitian ini berhasil menunjukkan pengoptimalan kinerja dengan peningkatan keakuratan yang signifikan antara sebelum dan sesudah *hyperparameter tuning*, mencegah *overfitting*, dan menaikkan efisiensi pelatihan.

Perbandingan optimalisasi hasil prediksi antara *Random Forest* dengan metode lainnya dan optimasi parameter *Grid Search* biasa dengan optimasi lainnya dituangkan

secara lengkap pada penelitian Iskandar & Ramdhani (2023). Penelitian ini menggunakan 5 algoritma yang dikomparasikan, yakni *Random Forest*, *Linear Regression, Neural Network, K-Nearest Neighbor,* dan *Decision Tree*. Selain itu, penelitian ini dilengkapi dengan komparasi antara optimasi parameter *Grid Search* dan *Evolutionary* dalam meningkatkan keakuratan hasil prediksi. Adapun data yang diaplikasikan pada penelitian ini adalah data deret waktu harga penutupan saham *Microsoft Corporation* (MSFT) periode 1 Maret 1986 hingga 25 Mei 2023. Algoritma *Random Forest* didapati menjadi algoritma dengan performa terbaik dalam memprediksi harga saham dibanding keempat algoritma lainnya dengan nilai RMSE sebesar 0,802, di mana kinerja paling baik ditunjukkan oleh *Random Forest* dengan rasio pembagian 0,8 dengan RMSE sebesar 0,777. Penggunaan optimasi parameter *Grid Search* pada *Random Forest* menunjukkan kinerja prediksi yang lebih akurat dibanding optimasi parameter *Evolutionary*, di mana nilai RMSE terendah masing-masing senilai 0,638 dan 0,651. Penelitian ini juga mengevaluasi hasil kinerja *Random Forest* tanpa optimasi dan dengan penambahan optimasi parameter *Grid Search* dalam melakukan analisis prediksi *time series* di mana ditemukan bahwa pengimplementasian optimasi *Grid Search* dapat diandalkan dalam mengoptimalkan hasil akurasi prediksi dan mampu memperkecil kesalahan prediksi semakin mendekati nilai aktual yang ditunjukkan oleh penurunan nilai RMSE dari sebelum ke sesudah optimasi.

Jange (2022) melakukan prediksi pada harga saham Bank BCA berdasarkan pada indikator teknikal dengan metode *XGBoost* dalam penelitiannya. Indikator teknikal yang digunakan diantaranya adalah *Simple Moving Average* (SMA), *Exponential Moving Average* (EMA), *Moving Average Convergence/Divergence* (MACD), dan *Relative Strength Index* (RSI) dengan total akhir atribut sebanyak 8, yakni SMA_5, SMA_10, SMA_15, SMA_30, EMA_9, MACD, MACD_Signal, dan RSI. Ditemukan bahwa EMA_9 menjadi variabel dengan tingkat kepentingan teratas dalam memprediksi harga saham Bank BCA melampaui indikator-indikator lainnya, diikuti variabel RSI, MACD, MACD_Signal, SMA_5, SMA_10, SMA_15, dan SMA_30.

Didapatkan hasil akurasi prediksi yang cukup baik dengan MAPE senilai 4,01% dengan penggunaan *Grid Search CV* dalam mencari parameter yang paling optimal.

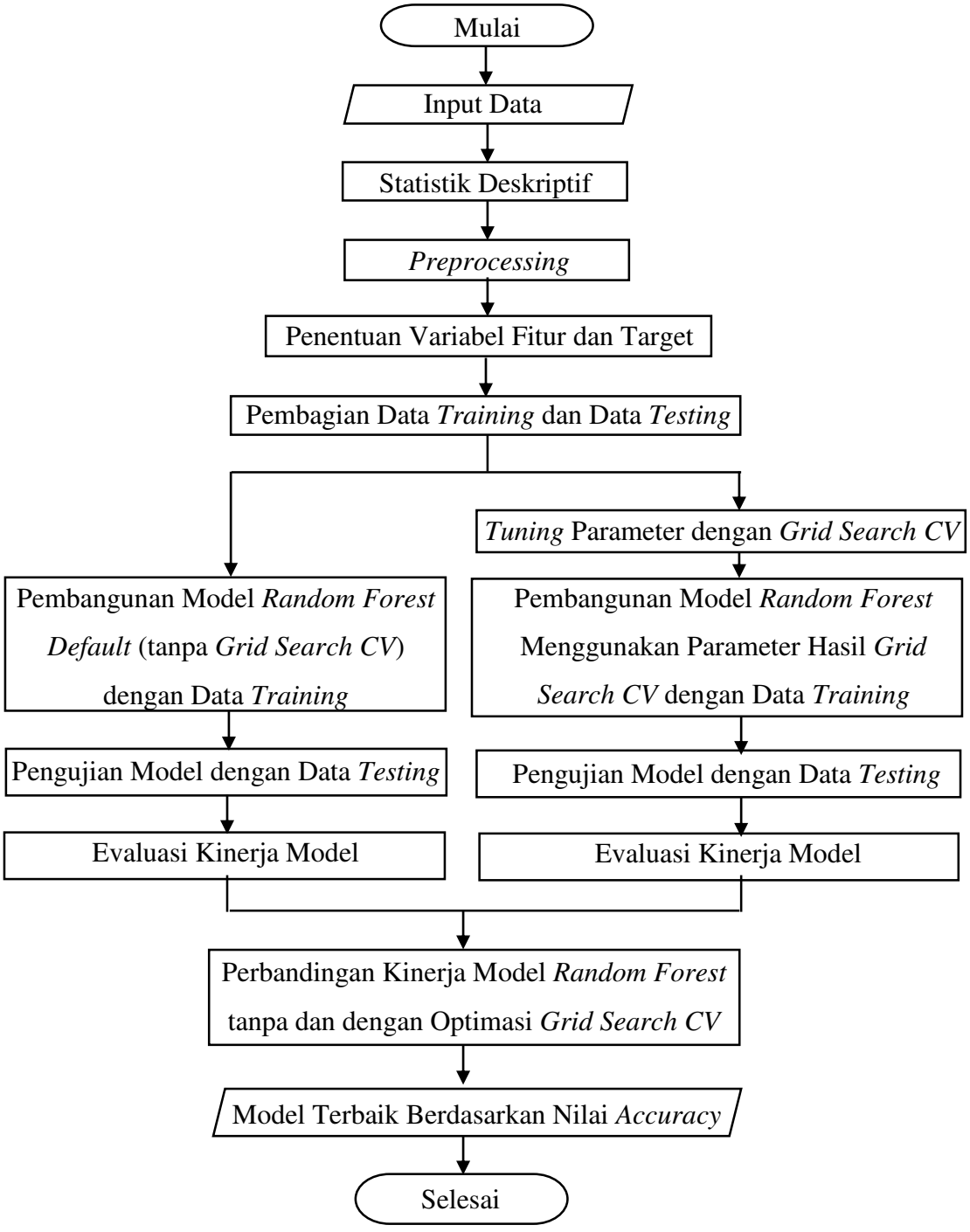
1.6 Metodologi Penelitian

Alur penelitian ini diawali dengan melakukan studi literatur pada bahan-bahan bacaan yang relevan dengan topik penelitian, khususnya seputar pengaplikasian *Data Mining* pada prediksi pergerakan harga saham beserta implementasi algoritma *Random Forest* dengan optimasi *Grid Search CV.* Data yang diaplikasikan adalah data sekunder berupa data historis harga saham INDF.JK yang didapat dari website *Yahoo Finance*.

Setelah data diperoleh dari sumber, dilakukan analisis statistik deskriptif untuk memberikan gambaran awal data. Langkah selanjutnya adalah melakukan *preprocessing* data, yaitu membersihan dan memastikan data siap untuk digunakan dalam proses pengolahan nantinya. Dilanjutkan dengan langkah penentuan variabel fitur dan target yang hendak dilibatkan pada proses prediksi. Pada tahap ini, dilakukan proses perhitungan nilai indikator teknikal yang didasarkan pada rumus dari tiap indikator. Proses dilanjutkan dengan pembagian data, yaitu membagi data menjadi data *training* dan data *testing* dengan proporsi yang diaplikasikan 80:20. Data *training* adalah data yang difungsikan untuk membangun model prediksi berdasarkan variabel yang ditetapkan sedangkan data *testing* merupakan data yang akan difungsikan untuk mengukur keakuratan model yang terbentuk.

Untuk melihat optimalitas *Grid Search CV* dalam meningkatkan performa *Random Forest*, maka dilakukan 2 alur kerja yang berbeda, yakni alur pertama memuat *Random Forest* tanpa *Grid Search CV* sedangkan alur kedua memuat *Random Forest* dengan optimasi *Grid Search CV*. Alur kerja pertama, *Random Forest* melakukan pembentukan model pada data dengan pengaturan parameter tanpa *tuning* parameter. Lain halnya dengan alur kerja kedua, optimasi *Grid Search CV* digunakan pada pengaturan parameter yang selanjutnya digunakan pada pembentukan model oleh *Random Forest*. Tiap model tersebut selanjutnya diuji keakuratan prediksinya dengan data *testing* dan dilihat tingkat kevalidan hasil prediksinya melalui ukuran evaluasi

kinerja model, serta dibandingkan hasilnya satu sama lain. Adapun tahapan yang diimplementasikan divisualisasikan pada *flowchart* penelitian berikut.



Gambar 1. 1 *Flowchart* Penelitian