

SKRIPSI

ANALISIS TINGKAT PENGEMBALIAN SAHAM-SAHAM
PADA INDEKS LQ45 DENGAN MODEL *RANDOM FOREST*
DI SEKTOR PERBANKAN



VINCENT RONALDI

NPM: 6161901011

PROGRAM STUDI MATEMATIKA
FAKULTAS TEKNOLOGI INFORMASI DAN SAINS
UNIVERSITAS KATOLIK PARAHYANGAN
2024

FINAL PROJECT

ANALYSIS OF STOCKS RETURN ON LQ45 STOCK INDEX
WITH *RANDOM FOREST* IN THE BANKING SECTOR



VINCENT RONALDI

NPM: 6161901011

DEPARTMENT OF MATHEMATICS
FACULTY OF INFORMATION TECHNOLOGY AND SCIENCES
PARAHYANGAN CATHOLIC UNIVERSITY
2024

LEMBAR PENGESAHAN

ANALISIS TINGKAT PENGEMBALIAN SAHAM-SAHAM PADA INDEKS LQ45 DENGAN MODEL *RANDOM FOREST* DI SEKTOR PERBANKAN

Vincent Ronaldi

NPM: 6161901011

Telah lulus ujian skripsi pada 22 Januari 2024 dengan penguji:
Maria Anastasia, M.Si., M.Act.Sc. dan Melania Eva Wulanningtyas, M.Pd.

Bandung, 31 Januari 2024

Menyetujui,

Pembimbing 1

Pembimbing 2

Felivia Kusnadi, M.Act.Sc.

Robyn Irawan, M.Sc.

Mengetahui,

Ketua Program Studi

Jonathan Hoseana, Ph.D.

PERNYATAAN

Dengan ini saya yang bertandatangan di bawah ini menyatakan bahwa skripsi dengan judul:

ANALISIS TINGKAT PENGEMBALIAN SAHAM-SAHAM PADA INDEKS LQ45 DENGAN MODEL *RANDOM FOREST* DI SEKTOR PERBANKAN

adalah benar-benar karya saya sendiri, dan saya tidak melakukan penjiplakan atau pengutipan dengan cara-cara yang tidak sesuai dengan etika keilmuan yang berlaku dalam masyarakat keilmuan.

Atas pernyataan ini, saya siap menanggung segala risiko dan sanksi yang dijatuhkan kepada saya, apabila di kemudian hari ditemukan adanya pelanggaran terhadap etika keilmuan dalam karya saya, atau jika ada tuntutan formal atau non-formal dari pihak lain berkaitan dengan keaslian karya saya ini.

Dinyatakan di Bandung,
31 Januari 2024



Vincent Ronaldi
NPM: 6161901011

ABSTRAK

Dalam dunia keuangan modern, sudah tidak jarang perusahaan menerbitkan saham untuk mengumpulkan modal dari *trader*. Dari pihak pembeli *trader*, saham menawarkan potensi untuk menumbuhkan kekayaan individu, namun membeli saham juga melibatkan risiko dan ketidakpastian, yakni turunnya harga saham saat mempertimbangkan untuk dijual dibandingkan harga saham saat dibeli. Dalam skripsi ini, algoritma *random forest* digunakan untuk melakukan prediksi tingkat pengembalian saham yang dapat membantu *trader* untuk membuat keputusan investasi guna memitigasi risiko. *Random forest* adalah pengembangan dari algoritma *regression tree* untuk mengurangi variansi dan meningkatkan akurasi hasil prediksi. Model *random forest* digunakan untuk memprediksi tingkat pengembalian saham dari data bulanan indeks saham LQ45 di sektor perbankan selama awal bulan Februari 2019 hingga akhir bulan Januari 2023, kemudian menentukan saham yang terbaik untuk diinvestasikan. Alasan digunakan saham dari data bulanan indeks saham LQ45 di sektor perbankan adalah adanya peran sebagai indikator ekonomis di Indonesia dan banyaknya peminatan *trader* bahkan dari luar negara dalam saham perbankan. Dapat dievaluasi plot longitudinal tingkat pengembalian saham dan hasil prediksi interval tingkat pengembalian saham individu setiap saham dalam indeks LQ45 di sektor perbankan. Berdasarkan observasi hasil penelitian, model *random forest* cukup baik dalam memprediksi tingkat pengembalian saham dengan hasil prediksi nilai tingkat pengembalian saham setiap saham masih sering berada dalam 95% *confidence interval* hasil prediksi dengan pengecualian saham ARTO yang berfluktuasi jauh di bawah nilai ekspektasi dan batas bawah *confidence interval*. Saham dengan potensi investasi terbaik dalam indeks LQ45 di sektor perbankan adalah BBCA dan BMRI.

Kata-kata kunci: *regression tree*; *random forest*; tingkat pengembalian saham; indeks saham LQ45.

ABSTRACT

In the modern finance world, it is not uncommon for companies to issue stock or shares to raise fund from traders. From the trader side, stock offer potential to grow individual wealth, but buying stock also involves risk and uncertainty. The most common risk and uncertainty faced by traders when buying stock is a decrease in the price of stock when considering to sell the stock compared to the initial price of stock when buying the stock. In this thesis, random forest algorithm will be used to make predictions of stock return that can help make good investment decisions so as to reduce these risk and uncertainty. Random forest is an improvement or development of regression tree algorithm to reduce variance and improve the accuracy of prediction results. Random forest model will be used to predict monthly stock return from the monthly data of the LQ45 stock index in the banking sector in a period of 4 years, namely from the beginning of February 2019 to the end of January 2023, then choose the best stock to invest. Reasons for using monthly data of the LQ45 stock index in the banking sector include its role as an economic indicator in Indonesia and the high interest from traders, even from foreign countries, in banking stocks. The longitudinal plot of monthly stock return and the prediction results of individual monthly stock return interval for each stock in the LQ45 index in the banking sector can be evaluated. Based on the observation from research result, random forest model is pretty good in predicting monthly stock return with the predicted results of the monthly stock return for each stock often fall within the 95% confidence interval of the prediction results, with the exception of ARTO stock, which fluctuates significantly below the expected value and the lower limit of the confidence interval. Some stocks with the best investment potential in the LQ45 index in the banking sector are BBKA dan BMRI.

Keywords: regression tree; random forest; stock return; LQ45 stock index.

Waktu yang diberikan pada diri sendiri adalah benih kebijaksanaan yang akan tumbuh menjadi pohon kesuksesan.



KATA PENGANTAR

Puji dan syukur yang mendalam kepada Tuhan Yang Maha Esa, serta menghaturkan salam dan penghormatan kepada Sang Buddha, sebagai penuntun spiritualitas dan sumber ilham dalam perjalanan kehidupan ini, penulis dapat memulai, mengerjakan, dan menyelesaikan skripsi dengan judul "Analisis Tingkat Pengembalian Saham-Saham pada Indeks LQ45 dengan Model *Random Forest* di Sektor Perbankan" sebagai salah satu syarat untuk menyelesaikan studi di Program Studi Matematika, Fakultas Teknologi Informasi dan Sains, Universitas Katolik Parahyangan. Penulis berharap skripsi ini dapat bermanfaat bagi para pembaca.

Pada kesempatan kali ini, penulis ingin mengucapkan kepada pihak-pihak yang telah memberikan dukungan serta bantuan secara langsung maupun tidak langsung, yaitu:

1. Papa, Mama, Cindy, Viondy, Verrick, dan Pupi karena sudah menjadi keluarga yang selalu memberikan dukungan dan semangat kepada penulis selama ini termasuk proses penyusunan skripsi ini.
2. Ibu Felivia Kusnadi, M.Act.Sc. selaku dosen pembimbing 1 yang sudah bersedia meluangkan waktunya untuk memberikan ilmu, bimbingan, saran, arahan, dan dukungan untuk penulis selama proses penyusunan skripsi ini.
3. Bapak Robyn Irawan, M.Sc. selaku dosen pembimbing 2 yang sudah bersedia meluangkan waktunya untuk memberikan ilmu, bimbingan, saran, arahan, dan dukungan untuk penulis selama proses penyusunan skripsi ini.
4. Ibu Maria Anastasia, M.Si., M.Act.Sc. selaku ketua tim penguji yang telah memberikan saran beserta kritik untuk skripsi ini.
5. Ibu Melania Eva Wullaningtyas, M.Pd. selaku anggota tim penguji yang telah memberikan saran beserta kritik untuk skripsi ini.
6. Bapak Dr. Daniel Salim selaku koordinator skripsi yang telah memberikan arahan untuk penulis sehingga penulis dapat menyelesaikan skripsi ini tepat pada waktunya.
7. Marco, Michael, Desmond, Ko Sandhi, Hernando, Ko Willy, Louis, Ci Richelle, Ko Frengki, dan Mervin karena sudah menjadi teman yang selalu menghibur, memberi dukungan, dan bantuan selama perkuliahan.
8. Diovensen dan Hermanto karena sudah menjadi teman terbaik semenjak masa sekolah yang selalu menghibur, memberi dukungan, dan bantuan selama perkuliahan.
9. Teman-teman matematika 2019 yang selalu mendukung dan membantu satu sama lain beserta memberikan pengalaman hidup selama perkuliahan.
10. Seluruh dosen dan staf Tata Usaha FTIS yang telah memberikan ilmu, waktu, dan bantuan kepada penulis selama perkuliahan.
11. Seluruh pihak lainnya yang tidak dapat disebutkan satu per satu atas segala dukungan dan bantuannya.

Penulis berharap adanya saran dan kritik dari para pembaca yang dapat membantu skripsi ini untuk menjadi lebih baik lagi.

Bandung, 31 Januari 2024

Penulis



DAFTAR ISI

KATA PENGANTAR	viii
DAFTAR ISI	x
DAFTAR GAMBAR	xii
DAFTAR TABEL	xiv
DAFTAR ALGORITMA	xv
DAFTAR NOTASI	xvi
1 PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	2
1.3 Tujuan	3
1.4 <i>State of the Art</i>	3
1.5 Batasan Masalah	4
2 LANDASAN TEORI	6
2.1 Indeks Saham LQ45 dan Metrik Analisis Saham	6
2.2 <i>Machine Learning</i>	9
2.3 <i>Regression Tree</i>	9
2.3.1 Contoh Perhitungan <i>Split Value Root Node</i>	12
2.4 <i>Random Forest</i>	13
2.5 <i>Goodness-of-fit</i>	14
2.5.1 <i>Mean Squared Error & Root Mean Squared Error</i>	14
2.5.2 <i>Percent of Variation Explained</i>	16
3 METODOLOGI PENGOLAHAN DATA	17
3.1 Deskripsi dan Metode Pengumpulan Data	17
3.2 Pembentukan Model-Model <i>Random Forest</i>	23
3.2.1 Percobaan 1	23
3.2.2 Percobaan 2	24
3.2.3 Percobaan 3	24
3.2.4 Percobaan 4	24
3.2.5 Percobaan 5	24
4 HASIL PENELITIAN DAN ANALISIS	25
4.1 Pengecualian Fitur pada Percobaan 4	26
4.2 Hasil <i>Goodness-of-Fit</i> Semua Percobaan	28
4.3 Hasil Prediksi Interval <i>Monthly Stock Return</i>	31
5 KESIMPULAN DAN SARAN	37

5.1 Kesimpulan	37
5.2 Saran	37
DAFTAR REFERENSI	39
A HASIL EKSPERIMEN	40



DAFTAR GAMBAR

2.1	Struktur <i>decision tree</i>	10
4.1	Longitudinal <i>monthly stock return</i> gabungan setiap saham dalam indeks LQ45 di sektor perbankan	25
4.2	Longitudinal <i>monthly stock return</i> individu setiap saham dalam indeks LQ45 di sektor perbankan	26
4.3	<i>Variable importance</i> percobaan 3	27
4.4	Distribusi dari <i>minimal depth</i> percobaan 3	28
4.5	<i>MSE OOB</i> dan <i>MSE testing</i> untuk semua kemungkinan banyaknya prediktor m_{try} pada model <i>random forest</i> percobaan 4	30
4.6	<i>MSE OOB</i> untuk semua kemungkinan banyaknya <i>bagged tree</i> dibangun pada model <i>random forest</i> percobaan 5	31
4.7	<i>Variable importance</i> percobaan 5	32
4.8	Distribusi dari <i>minimal depth</i> percobaan 5	32
4.9	<i>MSE OOB</i> dan <i>MSE testing</i> untuk semua kemungkinan banyaknya prediktor m_{try} pada model <i>random forest</i> percobaan 5	33
4.10	Longitudinal <i>monthly stock return</i> dan hasil prediksi interval <i>monthly stock return</i> individu setiap saham dalam indeks LQ45 di sektor perbankan percobaan 5 bagian 1	34
4.11	Longitudinal <i>monthly stock return</i> dan hasil prediksi interval <i>monthly stock return</i> individu setiap saham dalam indeks LQ45 di sektor perbankan percobaan 5 bagian 2	34
A.1	Longitudinal <i>monthly stock return</i> dan hasil prediksi interval <i>monthly stock return</i> individu setiap saham dalam indeks LQ45 di sektor perbankan percobaan 1 bagian 1	40
A.2	Longitudinal <i>monthly stock return</i> dan hasil prediksi interval <i>monthly stock return</i> individu setiap saham dalam indeks LQ45 di sektor perbankan percobaan 1 bagian 2	41
A.3	<i>MSE OOB</i> untuk semua kemungkinan banyaknya <i>bagged tree</i> dibangun pada model <i>random forest</i> percobaan 1	41
A.4	<i>Variable importance</i> percobaan 1	42
A.5	Distribusi dari <i>minimal depth</i> percobaan 1	42
A.6	<i>MSE OOB</i> dan <i>MSE testing</i> untuk semua kemungkinan banyaknya prediktor m_{try} pada model <i>random forest</i> percobaan 1	43
A.7	Longitudinal <i>monthly stock return</i> dan hasil prediksi interval <i>monthly stock return</i> individu setiap saham dalam indeks LQ45 di sektor perbankan percobaan 2 bagian 1	43
A.8	Longitudinal <i>monthly stock return</i> dan hasil prediksi interval <i>monthly stock return</i> individu setiap saham dalam indeks LQ45 di sektor perbankan percobaan 2 bagian 2	44
A.9	<i>MSE OOB</i> untuk semua kemungkinan banyaknya <i>bagged tree</i> dibangun pada model <i>random forest</i> percobaan 2	44
A.10	<i>Variable importance</i> percobaan 2	45
A.11	Distribusi dari <i>minimal depth</i> percobaan 2	45
A.12	<i>MSE OOB</i> dan <i>MSE testing</i> untuk semua kemungkinan banyaknya prediktor m_{try} pada model <i>random forest</i> percobaan 2	46
A.13	Longitudinal <i>monthly stock return</i> dan hasil prediksi interval <i>monthly stock return</i> individu setiap saham dalam indeks LQ45 di sektor perbankan percobaan 3 bagian 1	46

A.14 Longitudinal <i>monthly stock return</i> dan hasil prediksi interval <i>monthly stock return</i> individu setiap saham dalam indeks LQ45 di sektor perbankan percobaan 3 bagian 2	47
A.15 <i>MSE OOB</i> untuk semua kemungkinan banyaknya <i>bagged tree</i> dibangun pada model <i>random forest</i> percobaan 3	47
A.16 <i>MSE OOB</i> dan <i>MSE testing</i> untuk semua kemungkinan banyaknya prediktor m_{try} pada model <i>random forest</i> percobaan 3	48
A.17 Longitudinal <i>monthly stock return</i> dan hasil prediksi interval <i>monthly stock return</i> individu setiap saham dalam indeks LQ45 di sektor perbankan percobaan 4 bagian 1	48
A.18 Longitudinal <i>monthly stock return</i> dan hasil prediksi interval <i>monthly stock return</i> individu setiap saham dalam indeks LQ45 di sektor perbankan percobaan 4 bagian 2	49
A.19 <i>MSE OOB</i> untuk semua kemungkinan banyaknya <i>bagged tree</i> dibangun pada model <i>random forest</i> percobaan 4	49
A.20 <i>Variable importance</i> percobaan 4	50
A.21 Distribusi dari <i>minimal depth</i> percobaan 4	50



DAFTAR TABEL

2.1	Tabel contoh data S	12
2.2	Tabel hubungan terikat antara fitur dosis dan tingkat efektif terurut dari dosis terkecil sampai terbesar	12
2.3	Tabel SSE dosis	13
2.4	Tabel SSE usia	13
3.1	Tabel <i>stock price</i>	18
3.2	Tabel <i>stock trading</i>	18
3.3	Tabel <i>financial data and ratio</i>	18
3.4	Tabel historis harian BBKA	19
3.5	Tabel <i>monthly stock return</i> BBKA	21
3.6	Tabel perubahan harga saham BBKA	22
4.1	Tabel <i>goodness-of-fit</i> semua percobaan	28
4.2	Tabel <i>percent of variation explained</i> untuk kombinasi <i>hyperparameter tuning</i> percobaan 5	30
4.3	Tabel kategori setiap saham	35
4.4	Tabel prediksi <i>monthly stock return</i> Februari 2023	36

DAFTAR ALGORITMA

1	Algoritma <i>regression tree</i>	11
2	Algoritma <i>random forest</i> untuk kasus regresi	15



DAFTAR NOTASI

BM_i	<i>book to market value</i> bulan ke- i
MSE	ukuran dari rata-rata perbedaan kuadrat antara nilai prediksi dan nilai sebenarnya
$Momentum_{k\text{ look-back}}$	momentum saham untuk k <i>month look-back period</i>
P	banyaknya fitur prediktor
RET_i	<i>monthly stock return</i> bulan ke- i
$RMSE$	akar kuadrat dari MSE
$RSI_{k\text{ look-back}}$	<i>relative strength index</i> untuk k <i>month look-back period</i>
SSE	penjumlahan dari jarak error masing data dengan rata rata data kelompok
S_1	kumpulan data kelompok 1 setelah pemisahan dari data awal
S_2	kumpulan data kelompok 2 setelah pemisahan dari data awal
$\% Var Explained$	ukuran akurasi atau kekuatan prediksi dari model dengan membandingkan variansi dari hasil prediksi dengan variansi dari fitur <i>output</i> data asli
m_{try}	banyaknya prediktor yang dipilih acak sebagai <i>subset</i> dipertimbangkan pada setiap pemisahan cabang <i>bagged tree</i>
n_{tree}	banyaknya <i>decision tree</i> yang dibangun di <i>random forest</i>
$nodesize$	parameter penentuan pemisahan <i>decision node</i>

BAB 1

PENDAHULUAN

1.1 Latar Belakang

Saham adalah surat bukti yang menunjukkan kepemilikan sebagian aset suatu perusahaan yang memberikan hak atas pembagian dividen dan keuntungan lainnya sesuai dengan besar aset yang disetor ¹. Dalam dunia keuangan modern, sudah tidak jarang perusahaan menerbitkan saham untuk mengumpulkan modal dari *trader*. Modal yang dikumpulkan ini kemudian digunakan oleh perusahaan untuk pengembangan dan pertumbuhan bisnis. Beberapa tipe *trader* saham yang umum, yaitu individu yang berinvestasi untuk kepentingan pribadi, institusi keuangan seperti perusahaan asuransi yang berinvestasi sebagai bagian dari strategi investasi perusahaan, dan lain-lain. Dari sisi *trader*, saham menawarkan potensi untuk menumbuhkan kekayaan dengan tujuan membantu individu dan institusi mencapai tujuan finansial. Tujuan finansial ini dapat berupa rencana pensiun di masa depan, mengatasi masalah inflasi di masa depan, dan lain-lain. Tujuan finansial tersebut menjadi faktor-faktor urgensi *trader* dalam membeli saham. Namun membeli saham juga melibatkan risiko dan ketidakpastian.

Salah satu risiko dan ketidakpastian yang bisa dihadapi oleh *trader* ketika membeli saham adalah peningkatan tingkat suku bunga yang menyebabkan turunnya harga saham mempertimbangkan untuk dijual dibandingkan harga awal saham saat dibeli. Apabila *trader* berinvestasi pada saham dari luar negeri, risiko yang bisa dihadapi yaitu fluktuasi nilai tukar dari mata uang yang dapat memengaruhi nilai investasi. Akibat dari berbagai risiko tersebut pada akhirnya secara tidak langsung menyatakan penurunan nilai portofolio dari *trader*. Untuk mengurangi risiko dan ketidakpastian ini, maka akan dilakukan prediksi tingkat pengembalian saham (*stock return*). *Stock return* menyatakan peningkatan atau penurunan harga saham dibandingkan harga awal saham ketika pertama dibeli yang dinyatakan dalam rasio atau persentase [1, hlm. 101].

Untuk memprediksi suatu nilai di masa yang akan datang dapat digunakan algoritma *machine learning*. Dengan meningkatnya ketersediaan data saham beserta kemajuan dalam algoritma *machine learning*, dapat digunakan salah satu algoritma untuk memprediksi tingkat pengembalian saham yaitu *random forest*. *Random forest* atau kadang disebut *decision tree forest* adalah algoritma yang mengombinasikan beberapa *decision tree* dengan prinsip dasar *bagging* yang menghasilkan data *bootstrap* dari data asli beserta *random feature selection* [2, hlm. 367]. *Random forest* dikenal kemampuannya untuk digunakan pada data dengan jumlah fitur prediktor yang sangat besar dengan akurasi lebih bagus dibandingkan iterasi sebelumnya yaitu *decision tree* dan *bagged tree* ([2, hlm. 367], [3, hlm. 198 & 199]). Prinsip *bagging* yang menghasilkan data *bootstrap* dilakukan dengan

¹<https://kbbi.web.id/saham> diakses pada 10 April 2023

menghasilkan *subset* data yang berukuran sama besar dengan melakukan pengambilan acak dari data asli dengan adanya pengembalian [3, hlm. 192]. Prinsip *bagging* ini mengasumsikan setiap observasi data saling bebas dan ketika setiap observasi data memiliki ketergantungan dasar seperti kasus data longitudinal terhadap waktu, maka asumsi saling bebas tersebut terlanggar [4, hlm. 285].

Breiman (2001) juga membuktikan *random forest* terlindungi dari masalah *overfitting* [3, hlm. 199]. Algoritma lain seperti *gradient boosting* dibandingkan *random forest* menawarkan kinerja prediktif yang sama baiknya tetapi komputasi pada *gradient boosting* lebih intensif sehingga cenderung membutuhkan waktu komputasi lebih lama dari *random forest* [3, hlm. 206]. Friedman juga mengakui bahwa *gradient boosting* hasil ciptaannya dapat memiliki kelemahan di mana terdapat masalah *overfitting* [3, hlm. 206]. Opsi algoritma lain yaitu *neural network* untuk kasus regresi memiliki kecenderungan terhadap masalah *overfitting* antara fitur prediktor dan *output* ([2, hlm. 227], [3, hlm. 143]). Kekurangan lain yaitu hasil dari *neural network* dinyatakan dalam model *black box* yang sulit atau hampir mustahil untuk diinterpretasi [2, hlm. 227].

Decision tree yang berupa iterasi sebelumnya dari *random forest* dapat diterapkan pada kasus klasifikasi dan kasus regresi. *Decision tree* untuk kasus klasifikasi (*classification tree*) digunakan ketika fitur *output* yang ingin diprediksi dari data dinyatakan dalam bentuk kategorikal sedangkan *decision tree* untuk kasus regresi (*regression tree*) digunakan ketika fitur *output* yang ingin diprediksi dari data dinyatakan dalam bentuk numerik. Sama kasusnya dengan *decision tree*, *random forest* yang berupa iterasi lanjutan dari *decision tree* juga dapat diterapkan pada kasus klasifikasi dan kasus regresi. Fleksibilitas pada model prediksi tergantung tipe fitur *output* data menjadi keunggulan *decision tree* dan *random forest*. Karena akan diprediksi *stock return* yang bernilai kontinu dan berbentuk numerik, maka akan digunakan *random forest* untuk kasus regresi.

Dalam penelitian ini, akan dibahas metode pengumpulan data beserta fitur yang akan digunakan dalam model *machine learning* untuk memprediksi *stock return*. Data yang digunakan adalah data dari indeks saham LQ45 di sektor perbankan yang dikumpulkan dari situs PT Bursa Efek Indonesia atau *Indonesian Stock Exchange (IDX)* dan *yahoo finance*. Digunakan data dari indeks saham LQ45 karena LQ45 mengukur kinerja harga dari saham-saham terbaik di pasar modal yang terdiri dari 45 emiten. Kemudian, digunakan data LQ45 dari sektor perbankan karena performansi saham perbankan dapat berperan sebagai indikator ekonomis di Indonesia seperti tingkat suku bunga, stabilitas keuangan, dan lain-lain. Alasan lain digunakan data LQ45 dari sektor perbankan adalah banyaknya peminatan *trader* bahkan dari luar negara dalam saham perbankan¹. Data akan diimplementasikan dalam algoritma *random forest* untuk kasus regresi. Selain itu, diterapkan metode dalam meningkatkan kinerja model *random forest* untuk kasus regresi pada data yang bersifat temporal atau ada efek waktu untuk membantu menangkap pola temporal atau waktu yang terdapat dalam data saham.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang sudah dijelaskan di atas, maka dapat ditentukan beberapa rumusan masalah sebagai berikut:

1. Bagaimana membentuk model prediksi tingkat pengembalian saham bulanan (*monthly stock*

¹<https://shorturl.at/JKPST> diakses pada 29 Januari 2024

- return*) pada data indeks saham LQ45 di sektor perbankan dengan menggunakan algoritma *random forest* untuk kasus regresi?
2. Bagaimana meningkatkan kinerja model prediksi tingkat pengembalian saham bulanan (*monthly stock return*) pada data indeks saham LQ45 di sektor perbankan dengan perubahan aturan pemisahan data, penambahan fitur prediktor, penghapusan fitur prediktor kurang penting, dan melakukan *hyperparameters tuning*?
 3. Bagaimana hasil evaluasi model prediksi tingkat pengembalian saham bulanan (*monthly stock return*) pada data indeks saham LQ45 di sektor perbankan berdasarkan hasil *goodness-of-fit* dengan metode *mean squared error*, *root mean squared error*, dan *percent of variation explained*?
 4. Apakah saham yang terbaik untuk diinvestasikan berdasarkan *plot* longitudinal *monthly stock return* dan hasil prediksi interval *monthly stock return* individu setiap saham dalam indeks LQ45 di sektor perbankan?

1.3 Tujuan

Tujuan dari penulisan skripsi ini adalah sebagai berikut:

1. Membentuk model prediksi tingkat pengembalian saham bulanan (*monthly stock return*) pada data indeks saham LQ45 di sektor perbankan dengan menggunakan algoritma *random forest* untuk kasus regresi.
2. Meningkatkan kinerja model prediksi tingkat pengembalian saham bulanan (*monthly stock return*) pada data indeks saham LQ45 di sektor perbankan dengan perubahan aturan pemisahan data, penambahan fitur prediktor, penghapusan fitur prediktor kurang penting, dan melakukan *hyperparameters tuning*.
3. Mengevaluasi model prediksi tingkat pengembalian saham bulanan (*monthly stock return*) pada data indeks saham LQ45 di sektor perbankan berdasarkan hasil *goodness-of-fit* dengan metode *mean squared error*, *root mean squared error*, dan *percent of variation explained*.
4. Menentukan saham yang terbaik untuk diinvestasikan berdasarkan *plot* longitudinal *monthly stock return* dan hasil prediksi interval *monthly stock return* individu setiap saham dalam indeks LQ45 di sektor perbankan.

1.4 State of the Art

Penelitian yang pernah dilakukan pada data saham menggunakan model *random forest* adalah penyeleksian saham untuk diinvestasikan dari pasar saham Tiongkok berdasarkan hasil *random forest* untuk kasus klasifikasi [5]. Penelitian tersebut melakukan penyeleksian saham dari saham yang masuk kategori *first class* yang berisi 20 saham dengan peluang tertinggi dipilih sebagai hasil dari *random forest* untuk kasus klasifikasi. Secara konsep, metode penyeleksian *random forest* untuk kasus klasifikasi tersebut menghitung suatu nilai peluang yang berbentuk numerik sebelum

dimasukkan dalam suatu kategori atau kelompok sebagai panduan investasi. Dapat diperiksa bagaimana apabila digunakan *random forest* untuk kasus regresi dengan hasil yang berbentuk numerik dan menunjukkan nilai prediksi fitur *output* secara langsung sebelum dimasukkan dalam suatu kategori atau kelompok sebagai panduan tambahan dalam pilihan investasi.

Penelitian lain yang pernah dilakukan pada data saham menggunakan model *random forest* adalah perbandingan kinerja antara dua versi dari model *linear regression*, *random forest* untuk kasus regresi, serta *gradient boosting* dalam prediksi tingkat pengembalian saham [6]. Penelitian tersebut bertujuan untuk menunjukkan bahwa apakah metode *machine learning* yang dapat menangkap pola atau hubungan kompleks antar fitur dan sering cocok digunakan untuk dataset dengan jumlah fitur yang besar memiliki kinerja yang lebih bagus dibandingkan model yang hanya mengasumsikan hubungan linear antara fitur prediktor dan *output* dalam prediksi tingkat pengembalian saham. Akan tetapi, penelitian tersebut tidak menunjukkan apakah model dengan kinerja terbaik tersebut juga memiliki hasil prediksi tingkat pengembalian saham yang akurat dengan data asli dan bagaimana memilih suatu saham unik sebagai potensi investasi dari model tersebut.

Penelitian lain yang pernah dilakukan pada data saham menggunakan model *random forest* khususnya untuk indeks LQ45 adalah perbandingan teknik *machine learning* dalam prediksi indeks LQ45 [7]. Penelitian tersebut bertujuan untuk menentukan model *machine learning* dengan kinerja terbaik dari 10 algoritma yang berbeda untuk kasus klasifikasi dalam prediksi indeks LQ45. Penelitian tersebut menyimpulkan bahwa *random forest* untuk kasus klasifikasi memiliki kinerja terbaik dalam prediksi indeks LQ45. Dapat diperiksa bagaimana apabila digunakan *random forest* untuk kasus regresi dalam prediksi tingkat pengembalian saham LQ45 dan mengelompokkannya ke dalam suatu kategori untuk menangkap sifat klasifikasi yang dinyatakan memiliki kinerja yang bagus dalam prediksi indeks LQ45.

Kontribusi yang diberikan dalam skripsi ini ialah metode *random forest* untuk kasus regresi diterapkan untuk memprediksi tingkat pengembalian saham dengan data indeks saham LQ45 di sektor perbankan. Selain itu, metode dalam meningkatkan kinerja model *random forest* untuk kasus regresi pada data yang bersifat temporal atau ada efek waktu untuk membantu menangkap pola temporal atau waktu yang terdapat dalam data saham. Kemudian, penyeleksian saham untuk diinvestasikan dari indeks saham LQ45 di sektor perbankan berdasarkan plot longitudinal *monthly stock return* dan hasil prediksi interval *monthly stock return* individu setiap saham dalam indeks LQ45 di sektor perbankan dari hasil *random forest* untuk kasus regresi. Ini dilakukan untuk tidak hanya berinvestasi pada saham dengan hasil prediksi *monthly stock return* yang mengalami kenaikan terbesar dibandingkan data asli *monthly stock return* 1 bulan sebelumnya, tetapi juga adanya konsiderasi hubungan antara hasil prediksi *monthly stock return* dan data asli *monthly stock return* untuk beberapa bulan sebelumnya sebagai panduan tambahan dalam pilihan investasi.

1.5 Batasan Masalah

Batasan masalah dalam penelitian ini adalah sebagai berikut:

1. Saham diasumsikan tidak ada dividen atau *capital gain*.
2. Pengumpulan data berdasarkan emiten saham yang masuk dalam indeks saham LQ45 di sektor perbankan pada periode perdagangan dari bulan Agustus 2022 sampai bulan Januari

2023 selama periode 4 tahun dari bulan Februari 2019 sampai bulan Januari 2023.

3. Data saham diambil dari situs PT Bursa Efek Indonesia atau *Indonesian Stock Exchange* (IDX) dan *yahoo finance*.

