

Prediksi *Return* Saham dengan Metode *Random Forest* dan Penerapannya untuk Seleksi Portofolio

1st Muthia Novi Syafira

Fakultas Informatika

Universitas Telkom

Bandung, Indonesia

muthianv@student.telkomuniversity.
ac.id

3rd Deni Saepudin

Fakultas Informatika

Universitas Telkom

Bandung, Indonesia

denisaepudin@telkomuniversity.ac.id

Abstrak— Portofolio dapat membantu investor untuk memilih saham mana yang terbaik dengan bobot yang optimal. Dalam berinvestasi investor tentu berharap mendapatkan keuntungan tinggi dengan resiko rendah. Pada optimasi portofolio indikator umum yang digunakan yaitu *expected return* dan resiko, namun mempertimbangkan dua hal itu saja tidak cukup untuk mendapatkan portofolio yang terbaik. Masalah yang dibahas pada tugas akhir ini yaitu seleksi portofolio menggunakan informasi dari prediksi *return* menggunakan metode *Random Forest*. Data yang digunakan adalah data indeks LQ45 selama 7 tahun (2015-2022). Hasil dari tugas akhir ini berupa data kelompok saham dari sektor yang berbeda. Pengujian dilakukan dengan menggunakan prediksi *return* dan tanpa prediksi *return* dengan melihat standar deviasi terkecil dan pertumbuhan portofolio terbesar. Kelima kelompok saham ini dilakukan seleksi portofolio dengan pembobotan *equal weight*. Hasil dari prediksi *return* menggunakan metode *Random Forest* dievaluasi menggunakan *Mean Absolute Percentage Error* (MAPE) mendapatkan nilai (23.62263) dan MAPE terkecil dari saham MNCN dengan nilai (9.90716). Dari hasil pengujian yang dilakukan portofolio dengan prediksi *return* menghasilkan standar deviasi lebih besar dan *average return* kecil, dibanding dengan portofolio tanpa prediksi *return* menghasilkan portofolio dengan standar deviasi lebih kecil dengan *average return* besar serta menghasilkan pertumbuhan portofolio lebih baik.

Kata Kunci — Portofolio, *Return*, LQ45, *Random Forest Regression*

I. PENDAHULUAN

A. Latar Belakang

Investasi pada aset keuangan kini semakin mengalami peningkatan yang cukup baik. Teknologi informasi juga mendukung sehingga mempermudah akses untuk berinvestasi [7]. Banyaknya investasi yang tersedia di pasar modal memberikan investor lebih banyak pilihan salah satunya adalah saham. Saham masih sering menjadi pilihan utama investor untuk melakukan investasi yang menguntungkan.

Risiko investasi saham terbilang paling besar dengan besarnya return yang diharapkan. Semakin besar return maka semakin besar pula resiko [1]. Dibutuhkan prediksi nilai saham beberapa tahun terakhir untuk melihat pergerakan pasar saham [10]. Oleh karena itu, sebelum investor memutuskan menginvestasikan dananya untuk berinvestasi sangat penting untuk seleksi portofolio saham yang optimal agar mendapatkan return maksimal.

Seleksi portofolio adalah seleksi yang menggunakan informasi dan teknik peramalan untuk mencari kinerja portofolio yang optimal [6]. Optimalisasi portofolio telah dipelajari sejak lama, yang paling umum digunakan yaitu teori portofolio modern Mean Variance [2]. Teori yang dipelopori oleh Markowitz (1952) ini mempertimbangkan nilai harapan data historis dan varian return data historis untuk membangun model optimalisasi portofolio [4]. Teori ini memiliki keterbatasan, seperti hipotesis terbatas serta kompleksitas komputasi untuk aset yang besar [3].

Seiring perkembangan teknologi yang semakin maju banyak peneliti yang telah mengembangkan teori ini seperti menggunakan informasi dari prediksi return. Pembelajaran mesin telah banyak menghasilkan model-model pembentukan portofolio pada pasar modal seperti random forest. Random forest banyak digunakan dalam memprediksi time series. Random forest merupakan kumpulan dari decision tree yang beroperasi menjadi suatu gabungan fungsional. Maka pada penelitian ini akan dibahas seleksi portofolio dengan mempertimbangkan prediksi return saham menggunakan metode random forest.

1. Topik dan Batasannya

Pada penelitian ini penulis menganalisis bagaimana menerapkan prediksi return saham menggunakan metode random forest, dan bagaimana membentuk portofolio saham yang mempertimbangkan prediksi return dengan menggunakan metode random forest. Pada penelitian ini terdapat beberapa batasan masalah, yaitu menggunakan data 6 saham LQ45 diperoleh dari finance.yahoo.com yang dipilih berdasarkan sektor yang berbeda. Data yang digunakan adalah harga penutupan (close) data mingguan (weekly) dalam kurun waktu selama 7 tahun (2015-2022).

2. Tujuan

Tujuan penelitian ini yaitu menerapkan metode random forest untuk memprediksi return saham, menghasilkan portofolio yang mempertimbangkan prediksi return dan tanpa prediksi return, serta mencari kinerja hasil portofolio saham terbaik.

II. KAJIAN TEORI

A. Studi Terkait

1. Return Saham

Return disebut sebagai keuntungan yang dihasilkan oleh investor dari investasi saham. Rumus yang digunakan untuk menghitung return sebagai berikut [5]:

$$R = \frac{P_t - P_{t-1}}{P_{t-1}} \tag{1}$$

Keterangan:

- = return saham pada waktu ke-t
- () = harga saham pada waktu ke-t
- (-1) = harga saham sebelum waktu ke-t

2. Portofolio

Portofolio adalah gabungan investasi yang dimiliki oleh investor. Setiap investor memiliki bobot dan porsi yang berbeda terhadap nilai portofolio. Bobot portofolio yang dilambangkan dengan (1, 2, 3, ...,). Ketika semua bobot dijumlahkan, jumlahnya adalah 1 atau 100%[5].

3. Return Portofolio

Return portofolio merupakan rata-rata tertimbang dari setiap rekuritas dalam portofolio. Secara sistematis return portofolio dapat dihitung sebagai berikut [6]:

$$R_p = \sum_{i=1}^n (w_i \cdot R_i) \tag{2}$$

Keterangan:

- R_p = return portofolio
- w_i = bobot dana saham i pada portofolio
- R_i = return actual saham i pada portofolio
- n = banyak saham pada portofolio

4. Average Return Portofolio

Average return portofolio digunakan untuk melihat pertumbuhan portofolio. Rumus average return portofolio sebagai berikut [5]:

$$\bar{X} = \frac{\sum_{i=1}^n R_p}{n} \tag{3}$$

Keterangan:

- \bar{X} = average return portofolio
- R_p = nilai return portofolio
- n = banyak data

5. Standar Deviasi Portofolio

Standar deviasi digunakan mencari persebaran data dan melihat seberapa dekat data dengan rata-rata dari sampel. Rumus standar deviasi sebagai berikut [5]:

$$\sigma = \sqrt{\frac{\sum_{i=1}^n (R_{p0} - \bar{x})^2}{n-1}} \tag{4}$$

Keterangan:

- σ = standar deviasi portofolio
- R_p = nilai return portofolio ke-i
- \bar{x} = average return portofolio
- n = banyak data

Semakin besar standar deviasi (σ) yang dihasilkan, semakin besar kemungkinan return actual berbeda dengan return harapan.

6. Equal Weight

Pembobotan equal weight adalah jenis pembobotan yang menetapkan nilai yang sama untuk semua saham dalam portofolio.

7. Random Forest Regression

Random Forest pertama kali dikembangkan oleh Leo Breiman. Random forest merupakan teknik pembelajaran mesin ensemble [10]. Model ini menghindari masalah overfitting karena selalu konvergen sehingga kelebihan random forest ini sering digunakan untuk prediksi saham [9]. Metode ini merupakan pengembangan dari decision tree dengan mengkombinasikan dari masing-masing tree yang terbaik menjadi multiple decision tree [8]. Metode random forest dapat diimplementasikan dengan dua metode prediksi yaitu klasifikasi dan regresi. Perbedaan klasifikasi dengan regresi adalah klasifikasi bekerja dengan cara memberi label pada setiap sampel. Sedangkan regresi digunakan untuk menemukan fungsi yang memodelkan data dengan meminimalkan kesalahan atau perbedaan antara nilai prediksi dan nilai sebenarnya.

Random forest bergantung pada sebuah nilai vector-vector random dengan distribusi yang sama pada semua tree yang masing masing decision tree memiliki kedalaman yang maksimal [4]. Secara umum, semakin banyak tree pada sebuah hutan (forest) maka semakin kuat juga hutan tersebut terlihat. Parameter yang diimplementasikan pada metode ini yaitu jumlah pohon keputusan ditetapkan menjadi 400, max_depth ditetapkan 70, min_sample_leaf ditetapkan 4. Tree dimulai dengan cara menghitung nilai entropy dan information gain berikut persamaan untuk menghitung nilai entropy dan information gain :

$$Entropy(Y) = - \sum p(c|Y) \log_2 p(c|Y) \tag{5}$$

Dimana Y adalah himpunan kasus dan (|Y) merupakan proporsi nilai Y terhadap kelas c.

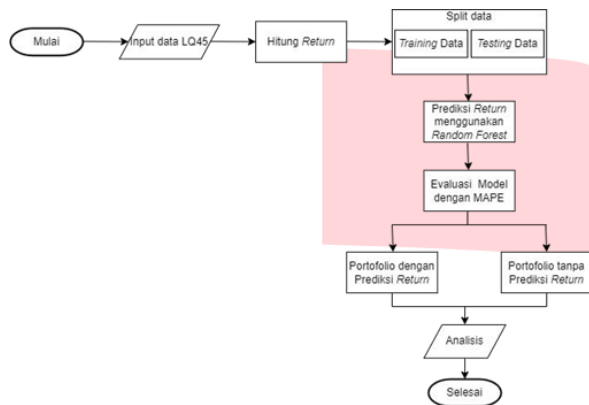
$$information\ gain(Y, a) = Entropy(Y) - \sum_{v \in Values(Y,a)} \frac{|Y_v|}{|Y|} Entropy(Y_v) \tag{6}$$

Dimana $Values(a)$ adalah semua nilai yang mungkin dalam himpunan kasus a , adalah subkelas dari Y dengan kelas v yang berhubungan dengan kelas a dan merupakan semua nilai yang sesuai dengan a .

III. METODE

A. Sistem yang Dibangun

Rancangan umum sistem ada penelitian ini yaitu menggunakan *flowchart* untuk menggambarkan alur kerja dari sistem yang dibangun dapat dilihat pada gambar dibawah ini.



GAMBAR 1
Perancangan Sistem yang dibangun

Dilihat pada Gambar 1 tahap pertama yang dilakukan adalah menginputkan data penutupan (data *close*) mingguan indeks LQ45 pada kurun waktu Januari 2015-Desember 2022. Tahap kedua adalah hitung *return actual* saham. Tahap ketiga adalah *split data training* dan *testing* untuk membangun model serta evaluasi model prediksi menggunakan MAPE. Tahap keempat memprediksi *return* saham dengan prediksi nilai harga yang sudah didapatkan menggunakan metode *random forest*. Tahap kelima adalah membentuk portofolio dengan prediksi *return* dan tanpa prediksi *return*. Dan tahap terakhir adalah melakukan analisis hasil dari tahapan sebelumnya.

B. Data

Data saham yang digunakan adalah data saham LQ45 yang diperoleh dari *finance.yahoo.com*. Data yang digunakan adalah data mingguan/*weekly* dari 6 sektor saham dalam kurun waktu 7 tahun (Januari 2015-Desember 2022). Berikut tabel data 6 sektor yang dipilih:

TABEL 1
Data LQ45

Kode	Nama Saham
BMRI	Bank Mandiri (Persero) Tbk
TLKM	Telekomunikasi Indonesia (Persero) Tbk
MNCN	Media Nusantara Citra Tbk
MECD	Medco Energi Internasional Tbk
HMSP	HM Sampoerna Tbk
ITMG	Indo Tambangraya Megah Tbk

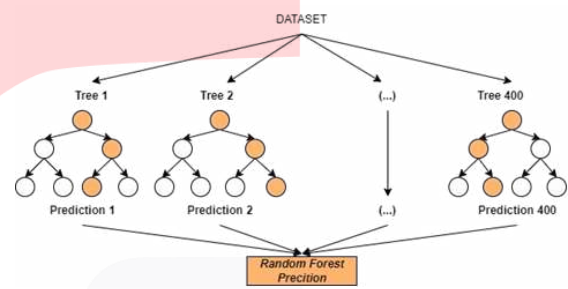
Untuk membentuk portofolio pada tugas akhir ini saham yang digunakan adalah saham dari sektor yang berbeda karena umumnya saham dari sektor yang sama memiliki pergerakan harga saham yang mirip. Apabila satu turun maka kecenderungan semua juga turun karena satu sektor. Maka data 6 saham pada tugas akhir ini dipilih berdasarkan sektor yang berbeda dari market cup yang paling tinggi.

C. Hitung Return

Dari data yang telah diinputkan maka akan dilakukan proses menghitung *return* saham menggunakan persamaan (1).

D. Split Data

Pada tahap ini dilakukan pemisahan data menjadi dua bagian yaitu *training data* digunakan untuk membangun model *random forest* dan *testing data* digunakan untuk menguji model setelah proses *training* selesai.



GAMBAR 2
Model Random Forest

Tahap selanjutnya yaitu prediksi *return* saham menggunakan *random forest*, berikut implementasi *random forest* yang dilakukan:

Membuat data sampel berukuran n . Gunakan sampel data untuk membangun *tree* ke- i ($i=1, 2, .. k$).

Membuat *decision tree* ke- i dari k data dengan dataset yang sudah dibangun sebelumnya. Dan ulangi kedua langkah tersebut sebanyak k .

Setelah mendapatkan *forest* (hutan) yang terdiri dari k *forest* acak selanjutnya lakukan prediksi nilai berdasarkan k buah *forest* tersebut. Jika sudah, maka cari rata-rata nilai terbanyak dari hasil prediksi setiap *tree*.

E. Evaluasi Model dengan MAPE

Pada tahap ini dilakukan perhitungan nilai MAPE dari *testing data* untuk evaluasi model prediksi yang telah terbentuk.

F. Portofolio dengan Prediksi Return

Setelah mendapatkan hasil prediksi *return* maka akan didapat saham-saham yang optimal. Dilakukan menggunakan pendekatan proporsi pembobotan equal weight, saham yang optimal akan dipilih berdasarkan dari besar atau kecil *return* prediksi yang dihasilkan dari setiap sektor saham. Pada metode equal weight portofolio saham-saham yang terpilih akan dijadikan bobot yang sama rata. Setelah itu dilakukan perhitungan *return* portofolio menggunakan persamaan (2).

G. Portofolio tanpa Prediksi Return

Setelah mendapatkan nilai return actual maka dilakukan pembobotan equal weight, sama seperti langkah sebelumnya yaitu saham yang terpilih akan dijadikan bobot yang sama rata. Setelah itu dilakukan perhitungan return portofolio menggunakan persamaan (2).

H. Analisis

Pada tahap terakhir dilakukan perhitungan standar deviasi dan average return portofolio. Standar deviasi dilakukan untuk melihat seberapa besar risiko pada suatu saham dan average return dilakukan untuk melihat pertumbuhan portofolio yang didapat. Parameter ini digunakan untuk melihat pertumbuhan dari portofolio yang terbentuk.

IV. HASIL DAN PEMBAHASAN

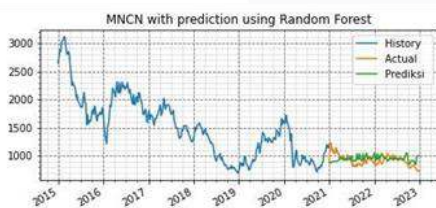
A. Evaluasi

1. Skenario Pengujian

- a. Data yang digunakan yaitu data 7 tahun LQ45 (01 Januari 2015 sampai 29 Desember 2022). Data training (01 Januari 2015 sampai 31 Desember 2020) dan Data test (07 Januari 2021 sampai 29 Desember 2022).
- b. Melakukan prediksi harga menggunakan random forest agar dapat melakukan prediksi return saham.
- c. Mencari bobot menggunakan equal weight.
- d. Mencari return portofolio dengan persamaan (2), standar deviasi dengan persamaan (4) dan average return dengan persamaan (3) dari portofolio yang terbentuk.
- e. Membandingkan hasil analisis pertumbuhan portofolio, standar deviasi kecil dari setiap sektor portofolio.

2. Hasil Pengujian

Pada penelitian ini dilakukan prediksi harga saham menggunakan metode random forest. Gambar (2) menunjukkan hasil dari prediksi harga saham menggunakan metode random forest.



GAMBAR 3 saham MNCN.JK



GAMBAR 4 Saham HMSP.JK



GAMBAR 5 Saham TLKM.JK

Gambar 3 Hasil prediksi harga saham menggunakan random forest. Grafik yang berwarna biru menunjukkan data historis selama 6 tahun, grafik yang berwarna orange menunjukkan harga actual sedangkan grafik yang berwarna hijau menunjukkan hasil prediksi harga.

Hasil prediksi return menggunakan random forest dievaluasi menggunakan Mean Absolute Percentage Error (MAPE) dapat dilihat pada tabel 2.

TABEL 2 Hasil MAPE

Saham	MAPE
BMRI	17.93429
TLKM	13.41471
MNCN	9.90716
MEDC	24.24169
HMSP	11.64030
ITMG	64.59764

Dalam penelitian ini dilakukan seleksi portofolio dengan prediksi return menggunakan random forest dan seleksi portofolio tanpa prediksi return. Parameter yang diamati dalam penelitian ini yaitu standar deviasi, average return.

TABEL 3 sektor saham, standar deviasi, average return

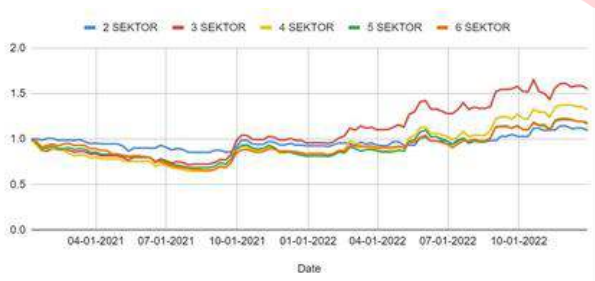
Sektor Saham	Dengan Prediksi Return		Tanpa Prediksi Return	
	Standar Deviasi	Average Return	Standar Deviasi	Average Return
BMRI, HMSP	0.02565	0.00120	0.02235	0.01664
BMRI, HMSP, ITMG	0.04005	0.00498	0.03156	0.02821
BMRI, HMSP, ITMG, MEDC	0.04078	0.00351	0.03187	0.03390
BMRI, HMSP, ITMG, MEDC, MNCN	0.03711	0.00215	0.02985	0.03401
BMRI, HMSP, ITMG, MEDC, MNCN, TLKM	0.03240	0.00209	0.02764	0.03521

Saham yang dipilih dikelompokkan menjadi beberapa sektor. BMRI, HMSP adalah 2 sektor, BMRI, HMSP, ITMG adalah 3 sektor, BMRI, HMSP, ITMG, MEDC adalah 4 sektor, BMRI, HMSP, ITMG, MEDC, MNCN adalah 5 sektor dan BMRI, HMSP, ITMG,

MEDC,MNCN,TLKM adalah 6 sektor. Nilai prediksi return yang telah didapat selanjutnya dilakukan strategi investasi yaitu membagi bobot sama besarnya disetiap sektor saham menggunakan equal weight. Setelah mendapatkan bobot dari setiap saham maka dihitung average return dengan persamaan (3) dan standar deviasi dengan persamaan (4) dari setiap sektor.

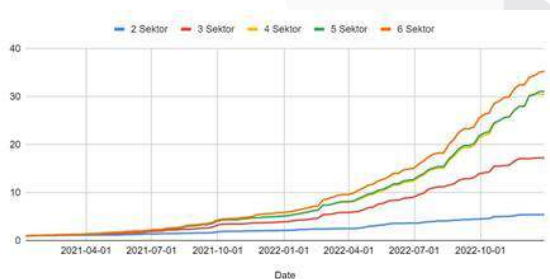
Dari tabel 3 menunjukan portofolio dengan prediksi return menghasilkan standar deviasi terkecil dimiliki oleh 2 sektor saham dengan nilai 0.02565 dan nilai average return dengan nilai 0.00120. Secara keseluruhan semakin bertambahnya jumlah saham average return turun seiring dengan standar deviasi yang naik turun, hal ini terjadi karena korelasi saham mendekati 0 maka standar deviasi atau resiko bisa mengecil seiring dengan dilakukannya diversifikasi.

Untuk portofolio tanpa prediksi return standar deviasi terkecil dimiliki oleh 6 sektor saham dengan nilai 0.02764 dan average return 0.03521. Semakin bertambahnya saham average return semakin naik dengan standar deviasi yang cenderung turun. Ini menunjukan hasil yang paling bagus dengan average return besar dan standar deviasi kecil.



GAMBAR 6
Pertumbuhan Portofolio dengan Prediksi Return

Pada gambar 6 dapat dilihat bahwa pertumbuhan portofolio dengan prediksi return menghasilkan pertumbuhan portofolio yang beragam dengan grafik yang naik turun dari setiap sektor, ini menunjukan pergerakan nilai return portofolio naik walaupun tidak signifikan.



GAMBAR 7
Pertumbuhan Portofolio tanpa Prediksi Return

Pada gambar 7 pertumbuhan portofolio tanpa return prediksi yang diuji dengan penambahan saham disetiap sektor menghasilkan pergerakan grafik yang lebih baik dengan nilai return portofolio yang semakin besar seiring dengan bertambahnya jumlah sektor saham.

V. KESIMPULAN

A. Kesimpulan

Kesimpulan yang didapatkan dalam tugas akhir prediksi return saham LQ45 dari 01 Januari 2015 sampai 29 Desember 2022 menggunakan metode *random forest* dengan jumlah pohon keputusan ditetapkan 400, *max_depth* ditetapkan 70 dan *min_sample_leaf* ditetapkan 4 menunjukan bahwa, kinerja portofolio dengan prediksi return menghasilkan pertumbuhan portofolio yang beragam, dengan nilai standar deviasi yang lebih besar dibanding dengan portofolio tanpa prediksi return yang menghasilkan portofolio terbaik dengan nilai standar deviasi lebih kecil serta pertumbuhan portofolio yang semakin besar seiring dengan bertambahnya jumlah sektor saham.

REFERENSI

- [1] Soeryana, E., Fadhlina, N., Rusyaman, E., & Supian, S. (2017, January). Mean-variance portfolio optimization by using time series approaches based on logarithmic utility function. In *IOP Conference Series: Materials Science and Engineering* (Vol.166, No. 1, p. 012003). IOP Publishing.
- [2] Bangun, D. H., Anantadjaya, S. P., & Lahindah, L. (2012). Portofolio Optimal Menurut Markowitz Model dan Single Index Model: Studi Kasus pada Indeks LQ45. *Journal of Management Studies*, 1(1), 70-93.
- [3] Ramadhan, R. D., Handayani, S. R., & Endang, M. G. W. (2014). Analisis pemilihan portofolio optimal dengan model dan pengembangan dari portofolio Markowitz (studi pada indeks BISNIS-27 di Bursa Efek Indonesia periode 2011-2013). *Jurnal Administrasi Bisnis*, 14(1).
- [4] Zhang, Y., Li, X., & Guo, S. (2018). Portfolio selection problems with Markowitz's mean-variance framework: a review of literature. *Fuzzy Optimization and Decision Making*, 17, 125-158.
- [5] Capinski, M., & Zastawniak, T. (2003). *Mathematics for finance. An Introduction*, 118-124.
- [6] Darusman, D., & Prasetyono, P. (2012). Analisis pengaruh firm size, book to market ratio, price earning ratio, dan momentum terhadap return portofolio saham. *Diponegoro Journal of Management*, 1(4), 212- 225.
- [7] Rifaldy, A., & Sedana, I. P. (2016). *Optimasi portofolio saham indeks bisnis 27 di bursa efek Indonesia (pendekatan model markowitz)* (Doctoral dissertation, Udayana University).
- [8] Saadah, S., & Salsabila, H. (2021). Prediksi Harga Bitcoin Menggunakan Metode Random Forest:(Studi Kasus: Data Acak Pada Masa Pandemi Covid-19). *Jurnal Komputer Terapan*, 7(1), 24- 32.

- [9] Ma, Y., Han, R., & Wang, W. (2021). Portfolio optimization with return prediction using deep learning and machine learning. *Expert Systems with Applications*, 165, 113973.
- [10] Vijn, M., Chandola, D., Tikkiwal, V. A., & Kumar, A. (2020). Stock closing price prediction using machine learning techniques. *Procedia computer science*, 167, 599-606.

