

# Analisis Kluster dalam Seleksi Aset dan Konstruksi dan Optimasi Portofolio Saham

Varid Vaya Yusuf<sup>1</sup>, Indwiarti<sup>2</sup>, Irma Palupi<sup>#3</sup>

<sup>1,2,3</sup> *Fakultas Informatika, Universitas Telkom  
Bandung, Indonesia*

<sup>1</sup> varidvaya@outlook.com

<sup>2</sup> indwiarti@telkomuniversity.ac.id

<sup>3</sup> irmapalupi@telkomuniversity.ac.id

# Corresponding author

## Abstract

Diversification in stock investment can be more optimal by grouping the stocks with the same characteristic. The hierarchical structure among the stocks exist in financial markets and can be analyzed by observing the correlation. Most of the previous studies only focused on the impact of cluster analysis on the portfolio's performance and rarely considered the assets selection in their benchmarks. This research proposes three alternative scenarios of asset selection for the cluster-based portfolio construction process as a new point of view in forming a portfolio construction benchmark. In its realization, the ward's method was utilized to clustering stocks based on in-sample data of 606 companies listed on IDX. Calinski-Harabasz index is applied to determine the number of clusters. The research was continued by portfolio construction with the tangency portfolio as the optimal portfolio preferences and three alternative scenarios of asset selection. Portfolio's performance measured by implementing both the Sharpe ratio and  $\Omega$  ratio on out-sample data. Cluster analysis conducted shows that groups of stocks formation have high quality. Moreover, the portfolio with cluster analysis also provides excellent performance, outperformed the portfolio without cluster analysis.

**Keywords:** assets selection, cluster analysis, portfolio diversification, ward's method.

## Abstrak

Diversifikasi investasi saham dapat dioptimalkan dengan klusterisasi saham-saham yang memiliki kesamaan karakteristik. Di dalam pasar keuangan, terdapat struktur hirarki antar saham yang saling berkorelasi. Penelitian ini mengajukan tiga skenario alternatif seleksi aset untuk proses konstruksi portofolio berbasis kluster sebagai sudut pandang baru dalam konstruksi portofolio. Secara teknis digunakan *Ward's method* untuk melakukan klusterisasi terhadap saham berdasarkan data *in-sample* dari 606 perusahaan tercatat di BEI. Indeks Calinski-Harabasz digunakan sebagai ukuran untuk menentukan jumlah kluster. Selanjutnya dilakukan konstruksi portofolio dengan *tangency portfolio* sebagai preferensi portofolio optimal dan seleksi aset dengan tiga skenario alternatif. performansi portofolio diukur menggunakan rasio Sharpe dan rasio  $\Omega$  terhadap data *out-sample*. Analisis kluster yang dilakukan menunjukkan kualitas yang luar biasa dalam kelompok-kelompok saham yang terbentuk. Portofolio dengan analisis kluster memberikan performansi yang sangat baik, melebihi portofolio tanpa analisis kluster.

**Kata Kunci:** diversifikasi, kluster, portofolio, saham, seleksi aset, *ward's method*.

## I. PENDAHULUAN

**D**IVERSIFIKASI Markowitz merupakan penemuan besar dalam bidang analisis portofolio investasi, dimana diversifikasi Markowitz berhasil menyadarkan investor bahwa korelasi antar aset sangat mempengaruhi kualitas diversifikasi investasi, dimana semakin kecil nilainya, semakin kecil risiko portofolionya. *Mean-Variance Optimization* (MVO) merupakan konsep yang pertama kali diperkenalkan pada tahun 1952 oleh Markowitz, dimana konstruksi portofolio berorientasi pada optimisasi *return* portofolio pada tingkat risiko tertentu atau optimasi risiko portofolio pada tingkat *return* tertentu [8].

Dalam teori finansial, manfaat diversifikasi dapat dioptimalkan dengan mengategorikan aset ke dalam kelas-kelas tertentu. Pernyataan tersebut didukung oleh paparan López de Prado dalam [2] yang menyatakan bahwa saham dapat digolongkan berdasarkan likuiditas, ukuran, sektor, dan daerahnya, dimana saham dalam satu kelas saling berkompetisi atas alokasinya. Namun dalam pengklasifikasian saham, sulit untuk mendefinisikan kelas-kelas saham secara jelas karena banyaknya variabel yang terlibat dalam sistem pasar saham. Mantegna pada tahun 1999 berhasil menemukan adanya struktur hirarki antar saham-saham di dalam pasar keuangan. Dalam tulisannya [7], beliau menjelaskan dengan baik bagaimana struktur hirarki antar saham dapat diterima sebagai alternatif dari klasifikasi saham sektoral dengan kemampuannya untuk merepresentasikan informasi dari sudut pandang ekonomi hanya dengan mengobservasi serangkaian harga saham yang saling berkorelasi.

Analisis portofolio dengan klusterisasi akhir-akhir ini mulai marak kembali dilakukan, seperti pada tahun 2010, Zhang dan Maringer dalam [12] berhasil menunjukkan dampak klusterisasi terhadap rasio Sharpe portofolio, dimana dilakukan klusterisasi terhadap saham-saham yang tercatat dalam pasar FTSE dengan rasio Sharpe dipandang sebagai fungsi objektifnya. Mereka mengajukan model alokasi aset berbasis kluster, dimana model tersebut memberikan nilai harapan *return* portofolio yang lebih tinggi dengan risiko portofolio yang sama dibandingkan dengan portofolio tanpa klusterisasi. Dewasa ini, pada tahun 2017 silam, León dkk, dalam [5] membandingkan enam portofolio hasil konstruksi oleh enam algoritma klusterisasi berbeda, baik klusterisasi partisi dan hirarki, dengan portofolio MVO tanpa analisis kluster dan melakukan *rebalancing* dalam interval waktu tertentu. León dkk, berkesimpulan bahwa portofolio dengan analisis kluster berhasil menunjukkan keseimbangan antara risiko dan *return* portofolio yang lebih baik dibandingkan portofolio MVO. Turut dijelaskan, hal tersebut terjadi karena alokasi aset tradisional tidak mampu mendeteksi adanya keterkaitan antar variabel yang kompleks dalam data. Tidak kalah penting, mereka juga memaparkan bahwa algoritma klusterisasi hirarki, terutama *ward's method* mampu melampaui performansi algoritma klusterisasi partisi dari perspektif performansi portofolio. Pada tahun yang sama, Tekin dan Gümüş pada [10] melakukan klusterisasi terhadap saham-saham yang tergabung ke dalam indeks Bursa Istanbul (BIST 100) berdasarkan sepuluh rasio finansial perusahaan. Hasil dari penelitiannya memaparkan bahwa klusterisasi terhadap saham menjadikannya dapat dipahami dan dianalisis secara lebih terukur sehingga memberikan sudut pandang tertentu dalam pengambilan keputusan investasi dan diversifikasi. Terlebih dalam [9], penerapan beberapa algoritma klusterisasi untuk diversifikasi saham menyimpulkan bahwa algoritma untuk tipe dataset yang *non-spherical*, yaitu agglomerative dan DBScan memberikan performansi diversifikasi yang baik.

Penelitian-penelitian terdahulu banyak berfokus pada dampak analisis kluster terhadap performansi portofolio, namun sedikit yang meninjau sisi seleksi aset dalam *benchmark*-nya. Fakta tersebut menjadi motivasi pelaksanaan penelitian ini dimana penulis mengajukan tiga skenario seleksi aset dalam proses konstruksi portofolio berbasis kluster sebagai sudut pandang baru dalam penyusunan *benchmark* konstruksi portofolio. Berdasarkan pemaparan León dkk, Mantegna, dan Ward, dalam penelitian ini, dipilih algoritma klusterisasi hirarki aglomeratif *ward's method* dalam melakukan analisis kluster karena terdapatnya struktur hirarki dalam pasar saham dan *ward's method* mampu memberikan performansi yang baik dalam kegiatan analisis kluster pada objek berupa saham. Sebagai alat ukur kemiripan saham, diimplementasikan algoritma klusterisasi berbasis korelasi, karena kemampuan luar biasa yang dimiliki korelasi untuk menggambarkan keterkaitan antar saham tanpa perlu melibatkan variabel-variabel dalam pasar saham yang kompleks. Selanjutnya dalam mengonstruksi portofolio optimal, dipilih *tangency portfolio* sebagai portofolio optimal karena kemampuannya untuk mengoptimalkan performansi portofolio berupa rasio Sharpe. Kemudian akan dipaparkan performansi portofolio yang merupakan hasil dari penelitian ini.

## II. METODE PENELITIAN

### A. Estimasi Parameter Konstruksi Portofolio

Dalam analisis portofolio, harapan *return* dan standar deviasi atau variansi *return* setiap aset, beserta kovariansi atau koefisien korelasi antar aset merupakan tiga parameter yang perlu didefinisikan terlebih dahulu sebelum melakukan konstruksi portofolio dengan memanfaatkan data historis harga saham. Sebelum dilakukan kalkulasi ketiga parameter tersebut, *return* aset perlu dihitung terlebih dahulu dengan rumus (1),

$$r_{it} = \frac{P_{it} - P_{it-1}}{P_{it-1}} \quad (1)$$

dimana  $r_{it}$  merupakan *return* aset ke- $i$  pada waktu ke- $t$  dan  $P_{it}$  adalah harga saham aset ke- $i$  pada waktu ke- $t$ . *Return* digunakan sebagai parameter untuk menghitung harapan *return* aset. Untuk menyederhanakan kalkulasi harapan *return* aset, *return* aset dianggap berdistribusi *uniform* dengan probabilitas  $1/T$  dimana  $T$  adalah banyaknya observasi historis yang dilakukan, sehingga harapan *return* aset dirumuskan dengan persamaan (2),

$$E(r_i) \approx \frac{1}{T} \sum_{t=1}^T r_{it} \quad (2)$$

dengan  $T$  merupakan ukuran sampel data historis.

Parameter terakhir yang perlu didefinisikan adalah koefisien korelasi antar aset- $i$  dan aset- $j$ , yang dalam hal ini digunakan koefisien korelasi Pearson  $\rho_{ij}$ , dihitung dengan formula (3).

$$\rho_{ij} = \frac{\sigma_{ij}}{\sigma_i \sigma_j} \quad (3)$$

$\sigma_{ij}$  merupakan nilai kovariansi sampel antar aset  $i$  dan  $j$ , sedangkan  $\sigma_i, \sigma_j$  berturut-turut merupakan nilai volatilitas sampel aset- $i$  dan aset- $j$ .

$$\sigma_{ij} = \frac{1}{T-1} \sum_{t=1}^T (r_{it} - E(r_i))(r_{jt} - E(r_j)) \quad (4)$$

### B. Ward's Method

*Ward's method* merupakan salah satu metode dalam algoritma klusterisasi hirarki aglomeratif yang direkomendasikan oleh Ward dalam [11] untuk menangani studi dengan jumlah objek yang besar  $n > 100$ , dimana sulit untuk mendefinisikan jumlah kluster optimal  $K$  secara presisi. *Ward's method* berbeda dari metode lainnya dimana minimalisasi *error sum of squares (ESS)* dipandang sebagai fungsi-objektifnya. *ESS* memiliki kemampuan untuk menunjukkan variasi dalam suatu kluster dan dirumuskan sebagai persamaan (5).

$$ESS = \sum_{i=1}^n (x_i - \bar{x})^2 \quad (5)$$

$x_i$  merupakan objek ke- $i$  dan  $\bar{x}$  merupakan rata-rata  $x$  keseluruhan objek, dengan fungsi objektif dapat dirumuskan seperti persamaan (6).

$$\min_{ESS} ESS_{total}(ESS) = \mathbf{1} \cdot ESS \quad (6)$$

dimana  $\mathbf{1}$  merupakan vektor satu berukuran  $1 \times k$  dengan  $k$  merupakan jumlah kluster, dan  $ESS$  merupakan vektor dari nilai *ESS* tiap kluster yang berukuran  $k \times 1$ .

Prosedur klusterisasi *Ward's method* secara umum dipaparkan dalam algoritma berikut.

- 1) **Inisiasi**, dimana jumlah kluster ( $k$ ) sama dengan jumlah objek ( $k = n$ ) dan mendefinisikan *ESS* sebagai vektor nol berukuran  $k \times 1$
- 2) **Selama**  $k \geq 1$ , **gabungkan** dua kluster dengan kriteria optimalisasi fungsi objektif sehingga  $k = k - 1$ , kemudian definisikan kembali vektor *ESS*.

### C. Algoritma Klusterisasi Berbasis Korelasi

Algoritma klusterisasi berbasis korelasi mengasumsikan kemiripan antara dua objek dapat diukur oleh koefisien korelasi. Dalam algoritma klusterisasi berbasis korelasi, nilai-nilai koefisien korelasi antar objek yang berjumlah  $n$  dimuat dalam matriks simetris definit positif berukuran  $n \times n$  dengan matriks diagonalnya bernilai satu. Dalam implementasinya, menurut Mantegna pada [7] dari hasil komunikasinya dengan Sornette, koefisien korelasi tidak dapat serta-merta digunakan sebagai ukuran jarak antar dua saham dikarenakan memiliki kodomain negatif. Sehingga perlu didefinisikan pemetaan jarak berdasarkan koefisien korelasi, seperti diformulasikan pada persamaan (7), dimana  $d_{ij}$  merepresentasikan jarak antara saham  $i$  dan saham  $j$ .

$$d_{ij} = \sqrt{2(1 - \rho_{ij})} \quad (7)$$

### D. Indeks Caliński-Harabasz

Tujuan dari analisis kluster adalah mengelompokkan objek dengan tingginya homogenitas objek dalam suatu kluster dan tingginya heterogenitas antar kluster. Dengan kata lain meminimalkan variansi antar objek di dalam tiap kluster ( $WGSS$ ) dan memaksimalkan variansi antar kluster ( $BGSS$ ). Nilai Caliński-Harabasz merujuk pada nilai *Variance Ratio Criterion* ( $VRC$ ) pada [1] yang merepresentasikan rasio antara  $BGSS$  terhadap  $WGSS$ .  $VRC$  dapat dirumuskan sebagai persamaan (8).

$$VRC = \frac{n - k}{k - 1} \times \frac{BGSS}{WGSS} \quad (8)$$

Dengan  $BGSS$  dan  $WGSS$  didefinisikan oleh persamaan berikut (9) dan (10).

$$BGSS = \sum_{i=1}^k n_i \|m_i - m\|^2 \quad (9)$$

$$WGSS = \sum_{i=1}^k \sum_{x \in c_i} \|x - m_i\|^2 \quad (10)$$

Dimana  $n_i$  merupakan jumlah objek di dalam kluster ke- $i$ ,  $m_i$  adalah *centroid* dari kluster ke- $i$ ,  $m$  merepresentasikan rata-rata keseluruhan data, dan  $c_i$  merupakan kluster ke- $i$ .

Nilai Caliński-Harabasz ini berkemampuan untuk menunjukkan kualitas kluster. Tingginya nilai Caliński-Harabasz menggambarkan kluster dengan tingginya homogenitas objek dalam suatu kluster dan tingginya heterogenitas antar kluster.

Dalam analisis kluster, jumlah kluster merupakan parameter yang esensial karena nilainya mempengaruhi keseimbangan antara kompresibilitas dan akurasi hasil analisis kluster. Disinilah nilai Caliński-Harabasz berperan, dimana dilakukan kalkulasi nilai Caliński-Harabasz untuk jumlah kluster yang berbeda-beda sebagai pembandingan untuk membantu pendefinisian jumlah kluster optimal. Namun perlu diingat, jumlah kluster dengan nilai Caliński-Harabasz tertinggi tidak selalu menjadi jumlah kluster optimal dalam analisis kluster, dimana dalam [1] dipaparkan dua pola pergerakan nilai Caliński-Harabasz.

- 1) Nilai Caliński-Harabasz berbanding terbalik dengan bertambahnya jumlah kluster, yang menandakan adanya kemungkinan struktur hirarki di dalam kluster.
- 2) Nilai Caliński-Harabasz sebanding dengan bertambahnya jumlah kluster, yang mengindikasikan bahwa kluster tidak terpisah dengan baik. Sering digunakan heuristik yang mengatakan besar kemungkinan jumlah kluster optimal ditemukan pada lokal optima atau titik balik dari pergerakan nilai Caliński-Harabasz.

### E. Silhouette Coefficient

Dalam analisis kluster, kualitas kluster yang dibentuk dapat dilihat melalui *silhouette coefficient* ( $SC$ ). Berpedoman pada buku [4] yang ditulis oleh Han dkk, perhitungan  $SC$  dapat dituliskan sebagai persamaan (11) berikut.

$$SC_i = \frac{b_i - a_i}{\max\{a_i, b_i\}} \quad (11)$$

Dimana  $SC_i$  merepresentasikan *silhouette coefficient* untuk objek ke- $i$ ,  $a_i$  merefleksikan rata-rata jarak antara objek ke- $i$  dengan semua objek di dalam kluster dimana objek ke- $i$  berada, sedangkan  $b_i$  merepresentasikan rata-rata jarak antara objek ke- $i$  dengan semua objek di kluster lain. Dalam analisis kluster dilakukan perhitungan  $SC$  untuk setiap objek.  $SC$  bergerak dalam rentang -1 hingga 1 ( $-1 \leq SC \leq 1$ ).  $SC$  bernilai positif jika  $a_i < b_i$ , yang mengindikasikan rapatnya kluster yang mengandung objek ke- $i$ , dan objek ke- $i$  terpisah jauh dengan kluster lain, sedangkan  $SC$  negatif menandakan objek ke- $i$  cenderung lebih dekat dengan objek dari kluster lain ketimbang dengan objek-objek di dalam klusternya sendiri, yang diakibatkan oleh  $a_i > b_i$ , dimana selalu diusahakan untuk meminimalisasi terjadinya kasus tersebut.

#### F. Tangency Portfolio

*Tangency portfolio* merupakan portofolio dimana *efficient frontier* bersinggungan dengan *capital market line* (CML). Dalam melakukan optimasi bobot dengan berpedoman pada buku [3] yang ditulis oleh Francis dan Kim, untuk konstruksi portofolio efisien yang mengandung aset berisiko sebanyak  $n$ , diformulasikan masalah optimasi dengan fungsi objektif (12) berikut.

$$\min_w L = \frac{1}{2} w^T \Sigma w + \lambda [E(r_p) - w^T E - (1 - w^T \mathbf{1}) r_f] \quad (12)$$

Dengan  $\lambda$  adalah pengali Lagrange,  $w$  adalah vektor bobot aset yang berukuran  $n \times 1$ ,  $\Sigma$  adalah matriks kovariansi aset berukuran  $n \times n$ ,  $E(r_p)$  merepresentasikan nilai harapan *return* portofolio,  $E$  merupakan matriks harapan *return* berukuran  $n \times 1$ , dan  $r_f$  adalah *return* aset bebas risiko. Untuk memperoleh solusi dari masalah optimasi, dilakukan turunan parsial untuk fungsi objektif terhadap dua variabel  $w$  dan  $\lambda$  yang dilakukan pada persamaan (13).

$$\begin{aligned} \frac{\delta L}{\delta w} &= \Sigma w - \lambda(E - r_f \mathbf{1}) = 0 \\ \frac{\delta L}{\delta \lambda} &= E(r_p) - w^T E - (1 - w^T \mathbf{1}) r_f = 0 \end{aligned} \quad (13)$$

Dari penurunan di atas, maka diperoleh dua persamaan, (14) dan (15).

$$w = \lambda \Sigma^{-1} (E - r_f \mathbf{1}) \quad (14)$$

$$E(r_p) = r_f + w^T (E - r_f \mathbf{1}) \quad (15)$$

Sehingga dari substitusi persamaan (14) ke persamaan (15), diperoleh nilai  $\lambda$  yang dapat dirumuskan sebagai persamaan (16).

$$\begin{aligned} \lambda &= \frac{[E(r_p) - r_f]}{H}, \text{ dengan} \\ H &= (E - r_f \mathbf{1})^T \Sigma^{-1} (E - r_f \mathbf{1}) \\ H &= B - 2r_f A + r_f^2 C \end{aligned} \quad (16)$$

Dimana konstanta  $A$ ,  $B$ , dan  $C$  terdefinisi di dalam persamaan (17) berikut.

$$\begin{aligned} A &= \mathbf{1}^T \Sigma^{-1} E \\ B &= E^T \Sigma^{-1} E \\ C &= \mathbf{1}^T \Sigma^{-1} \mathbf{1} \end{aligned} \quad (17)$$

Sebagai konklusi diperoleh bobot optimal untuk aset berisiko dan bebas risiko pada persamaan (18) dan (19).

$$w = \Sigma^{-1}(E - r_f \mathbf{1}) \left[ \frac{E(r_p) - r_f}{H} \right] \quad (18)$$

$$w_f = 1 - w^T \mathbf{1} \quad (19)$$

Dengan  $w_f$  merupakan bobot dari aset bebas risiko.

Bobot *tangency portfolio* didominasi oleh aset berisiko dengan bobot aset bebas risiko bernilai nol  $w_f = 0$ . Dari persamaan (17), (18), dan (19), bobot *tangency portfolio*  $w$  dapat didefinisikan sebagai persamaan.

$$w = \frac{\Sigma^{-1}(E - r_f \mathbf{1})}{A - r_f C} \quad (20)$$

### G. Analisis performansi Portofolio

Sebagai alat ukur performansi portofolio, digunakan nilai harapan *return*, variansi, rasio Sharpe, dan rasio  $\Omega$  dari masing-masing portofolio.

1) *Harapan return*: Nilai harapan *return* portofolio  $E(r_p)$  merupakan jumlah dari nilai harapan *return* tiap aset yang dikalikan dengan bobotnya, yang dapat dirumuskan sebagai persamaan (21) berikut.

$$E(r_p) = \sum_{i=1}^n w_i E(r_i) \quad (21)$$

Dengan  $w_i$  adalah bobot untuk aset ke- $i$ . Sesuai dengan tujuan dasar kegiatan investasi, nilai harapan *return* yang tinggi merupakan salah satu kriteria dasar dalam pemilihan portofolio.

2) *Variansi*: Variansi portofolio atau dalam interpretasinya dalam dunia investasi dikenal sebagai alat ukur volatilitas, merepresentasikan besarnya jarak naik-turunnya saham. Kalkulasi variansi portofolio  $\sigma_p^2$  dapat dilakukan dengan menggunakan persamaan (22).

$$\sigma_p^2 = \sum_{i=1}^n \sum_{j=1}^n w_i w_j \sigma_{ij} \quad (22)$$

Segala bentuk usaha investasi jelas bertujuan untuk memperoleh *return* yang maksimal dengan tingkat toleransi terhadap risiko tertentu yang tentunya akan diminimalkan. Sebagai contoh, jika investor dihadapkan oleh beberapa portofolio dengan *return* yang sama, tentu investor tersebut akan memilih portofolio dengan risiko atau variansi terendah.

3) *Rasio Sharpe*: Rasio Sharpe  $S_p$  merupakan ukuran untuk membandingkan performansi portofolio, yang merupakan perbandingan antara *excess return* ( $\bar{r}_p - r_f$ ) dengan standar deviasi portofolio  $\sigma_p$ . Ukuran ini dapat dihitung dengan formula (23)

$$S_p = \frac{\bar{r}_p - r_f}{\sigma_p}, \quad (23)$$

dimana  $\bar{r}_p$  adalah rata-rata *return* portofolio. Semakin tinggi nilai  $S_p$  mengindikasikan risiko yang ditanggung investor pantas/sesuai dengan *return* yang diperolehnya. Nilai  $S_p$  sendiri bias terhadap sampel. Telah dibuktikan secara matematika dalam [6], bahwa jika diasumsikan *return* bersifat *independent and identical distributed*, maka  $\frac{\bar{r}_p - r_f}{\sigma_p}$  akan menghampiri  $S_p$  populasi dengan akurat dan memenuhi selang berikut ini, dengan tingkat kepercayaan 95%.

$$S_p \text{ populasi} = S_p \pm 1.96 \times \sqrt{(1 + 0.5 S_p) / T}$$

4) *Rasio  $\Omega$* : Rasio  $\Omega$  mampu mengukur kemampuan suatu portofolio untuk mencapai suatu nilai *return* yang ditetapkan sebagai *threshold*  $r$ . Tingginya nilai Rasio  $\Omega$  menandakan performansi yang baik dari suatu portofolio, dimana portofolio berkemungkinan tinggi untuk mencapai, bahkan melampaui suatu nilai  $r$ . Dengan  $F(x)$  merupakan fungsi kumulatif padat peluang dari *return*  $x$ , rasio  $\Omega$  dengan peubah  $r$  dapat dirumuskan sebagai persamaan (24) berikut.

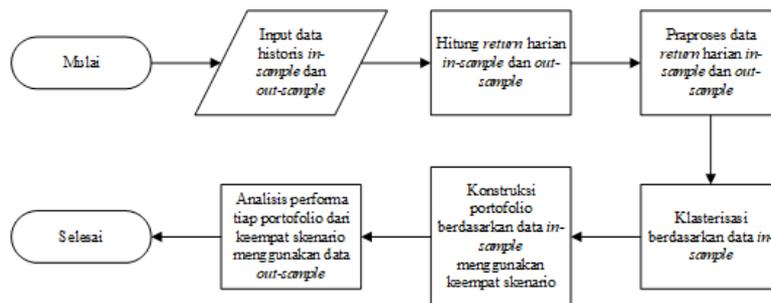
$$\Omega(r) = \frac{\int_r^\infty (1 - F(x))dx}{\int_{-\infty}^r F(x)dx} \approx \frac{\sum_{i=1}^T (g(x_i) - r)}{-\sum_{i=1}^T (h(x_i) - r)} \tag{24}$$

$$g(x) = \begin{cases} x, & x > r \\ r, & x \leq r \end{cases}, \quad h(x) = \begin{cases} r, & x \geq r \\ x, & x < r \end{cases}$$

Dalam penelitian ini, *return* portofolio berupa nilai diskrit *return* portofolio harian dan diasumsikan berdistribusi *uniform*, dengan  $x_i$  merupakan *return* pada hari ke- $i$ .

### III. IMPLEMENTASI ALGORITMA

Secara umum, implementasi dari algoritma disusun dalam sistem yang digambarkan oleh diagram alir pada Gambar 1 berikut.



Gambar 1: Diagram Alir Implementasi Algoritma.

#### A. Data

Data yang digunakan merupakan data historis *adjusted closing price* harian dari 606 perusahaan tercatat di BEI untuk masa observasi dari 2 Januari 2017 hingga 29 Desember 2017 sebagai data *in-sample* yang digunakan dalam proses klusterisasi dan konstruksi portofolio untuk mengestimasi parameter-parameter yang dibutuhkan. Sebagai data *out-sample*, digunakan data yang sama dengan masa observasi yang berbeda, yaitu dari 1 Januari 2018 hingga 31 Desember 2018 yang digunakan dalam proses analisis performansi portofolio untuk mengukur rasio Sharpe dan rasio  $\Omega$ . Dalam proses-proses selanjutnya, diperlukan data *return* harian *in-sample* dan *out-sample* tiap aset yang dapat dikalkulasi menggunakan data historis *adjusted closing price* harian dengan persamaan (1). Data yang diperoleh merupakan data yang konsisten, namun belum dapat dikatakan sebagai data berkualitas, dimana terdapat pencilan *return* saham dan hari tanpa transaksi yang merupakan *missing value*, sehingga perlu dilakukan praproses terhadap data yang dijelaskan pada subbab selanjutnya.

#### B. Praproses

Untuk memperoleh data yang berkualitas, dilakukan praproses pada data untuk mengatasi pencilan dan *missing value*. Untuk menangani *missing value*, pertama hari-hari tanpa transaksi dihapus, kemudian menghapus saham-saham dengan *missing value* dan pencilan. Sebagai hasil dari proses ini, tersisa 361 saham tanpa *missing value*.

C. *Estimasi Parameter Konstruksi Portofolio*

Setelah dilakukan praproses data, selanjutnya dilakukan estimasi nilai harapan *return* setiap aset, variansi *return* setiap aset, dan koefisien korelasi antar aset dari data *in-sample*, untuk selanjutnya dimanfaatkan sebagai parameter dalam klasterisasi dan konstruksi portofolio.

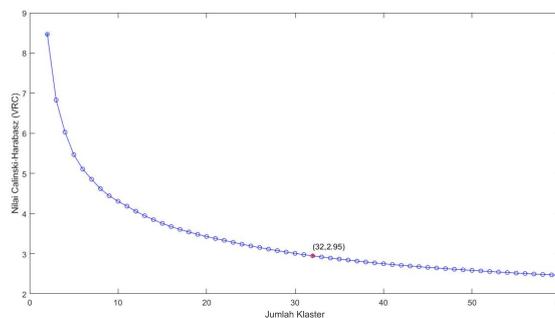
D. *Klasterisasi*

Klasterisasi yang dilakukan mengimplementasikan *Ward's method* berbasis nilai korelasi yang direpresentasikan dengan matriks  $D$  berukuran  $n \times n$  sebagai parameterinya. Setiap elemen matriks  $D$  dihitung dengan formula (7).

$$D = \begin{pmatrix} d_{11} & d_{12} & \dots & d_{1n} \\ d_{21} & d_{22} & \dots & d_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ d_{n1} & d_{n2} & \dots & d_{nn} \end{pmatrix}$$

Dalam pembentukan matriks  $D$ , perlu dilakukan *filtering* untuk saham-saham dengan nilai variansi *return* nol. Sehingga dengan dataset yang digunakan pada penelitian ini, tersisa hanya 357 saham dari total 606 saham.

Dalam proses klasterisasi, digunakan index Caliński-Harabasz sebagai acuan untuk menentukan jumlah kluster  $K$  optimal, yang mana nilai ini menyatakan rasio antara sebaran dalam kluster dan sebaran antar kluster. Pada Gambar 2, terlihat bahwa nilai Caliński-Harabasz tertinggi diperoleh dengan jumlah kluster  $K = 2$  dan seiring dengan bertambahnya jumlah kluster, nilai Caliński-Harabasz turut menurun dengan laju penurunan yang juga semakin mengecil. Pada saat jumlah kluster  $K = 32$ , yang ditandai dengan titik merah, laju penurunan nilai Caliński-Harabasz untuk pertama kali-nya mencapai nilai 1% dan memberikan nilai Caliński-Harabasz sebesar 2.95. Dengan penurunan yang tidak lagi signifikan, maka dianggap jumlah kluster  $k = 32$  sebagai jumlah kluster optimal.

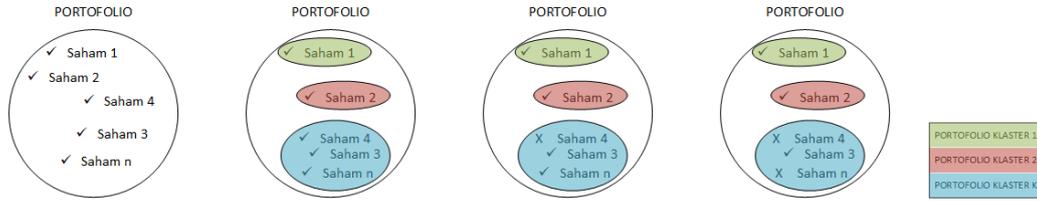


Gambar 2: Nilai Caliński-Harabasz.

Setelah mendefinisikan jumlah kluster optimal, selanjutnya dipilih hasil analisis kluster dengan  $K = 32$  sebagai kluster-kluster untuk melakukan seleksi aset dalam proses konstruksi portofolio, dimana kualitas kluster yang terbentuk dapat dilihat melalui *silhouette coefficient* dan matriks korelasi setelah pengurutan berdasarkan klasternya.

E. *Konstruksi Portofolio*

Pada penelitian ini, dilakukan Konstruksi portofolio dengan menerapkan empat skenario seleksi aset berbasis kluster yang dihasilkan. Konstruksi dengan *tangency portfolio* dianggap sebagai portofolio optimal berdasarkan keunggulan dari nilai tiga ukuran, yaitu ekspektasi *return*, variansi *return*, dan nilai rasio Sharpe dari portofolio. Bobot dibatasi agar tidak kurang dari 0.01%. Gambar 3 menggambarkan keempat skenario konstruksi tersebut. Tanda centang menandakan dilibatkan dalam konstruksi portofolio, sedangkan tanda silang berarti tidak.



(a) Ilustrasi Skenario 1 (b) Ilustrasi Skenario 2 (c) Ilustrasi Skenario 3 (d) Ilustrasi Skenario 4

Gambar 3: Ilustrasi Skenario Konstruksi Portofolio

Keempat skenario pada Gambar 3 mewakili kemungkinan aturan diversifikasi yang dilakukan setelah diperoleh hasil kluster dari saham-saham pada dataset. Skenario-skenario ini penting untuk pengujian performansi portofolio yang dihasilkan. Keempat skenario tersebut dapat dijabarkan sebagai berikut.

1) *Skenario 1:* Dari Gambar 3a dapat dilihat bahwa semua saham termasuk ke dalam portofolio dan tidak terdapat portofolio kluster. Dilakukan optimasi bobot terhadap semua saham tanpa melakukan klusterisasi, sehingga diperoleh vektor bobot  $w$  berukuran  $n \times 1$ .

2) *Skenario 2:* Ilustrasi portofolio skenario 2 ini dapat divisualisasikan oleh Gambar 3b. Terlihat bahwa Portofolio terdiri dari  $K$  portofolio kluster dimana setiap saham dalam kluster ikut berperan aktif memperebutkan alokasi. Secara algoritmis, dilakukan pembobotan di dalam tiap kluster, dimana tiap saham hanya bersaing untuk memperoleh alokasi dengan saham-saham yang ada di dalam kluster yang sama, dengan bobot maksimum tiap kluster adalah satu. Dari proses tersebut, dihasilkan matriks  $W$  berukuran  $n \times K$  yang berisi bobot saham dengan  $W_{ij}$  merepresentasikan bobot saham ke- $i$  dalam kluster  $C_j$  untuk  $j = 1, 2, \dots, K$  dan  $n = \max\{|C_j| \mid j = 1, 2, \dots, K\}$  menyatakan jumlah anggota kluster terbesar. Matriks bobot  $W$  dapat ditulis sebagai berikut.

$$W = \begin{pmatrix} W_{11} & W_{12} & \dots & W_{1K} \\ W_{21} & W_{22} & \dots & W_{2K} \\ \vdots & \vdots & \ddots & \vdots \\ W_{n1} & W_{n2} & \dots & W_{nK} \end{pmatrix}$$

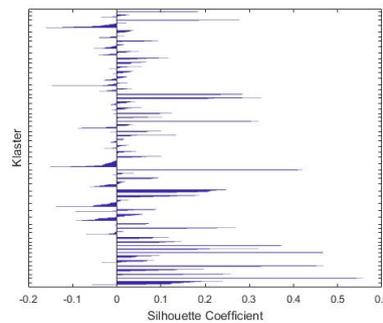
Selanjutnya portofolio dalam masing-masing kluster  $C_j, j = 1, 2, \dots, K$  dianggap sebagai satu aset untuk kemudian mengonstruksi alokasi portofolio antar kluster yang menghasilkan vektor bobot portofolio  $w_K$  berukuran  $K \times 1$  yang memenuhi  $(\mathbf{1}_K)^T w_K = 1$ . Sebagai tahap akhir dari skenario 2, dapat ditentukan bobot untuk masing-masing saham dalam portofolio utuh, yaitu  $W' = W \odot (\mathbf{1}_K (w_K)^T)$ . Matriks  $W'$  sudah pasti memenuhi  $(\mathbf{1}_n)^T W' \mathbf{1}_K = 1$ , dengan  $(\mathbf{1}_N) := [1 \ 1 \ \dots \ 1]_{(1 \times N)}^T$ .

3) *Skenario 3:* Seperti terlihat pada Gambar 3c, perbedaan skenario 3 dengan skenario 2 adalah tidak semua portofolio dalam kluster berperan aktif dalam alokasi. Saham-saham terpilih dari tiap kluster dinamakan sebagai saham representatif. Secara algoritma, dilakukan pembobotan di dalam tiap kluster, dimana tiap saham hanya bersaing untuk memperoleh alokasi dengan saham-saham yang ada di dalam kluster yang sama, dengan bobot maksimum tiap kluster adalah satu. Dari proses tersebut, dihasilkan matriks bobot  $W$  kemudian dipilih saham representatif yang memiliki bobot di atas nilai *threshold*. Tahap akhir dari skenario ini adalah melakukan optimasi bobot terhadap semua saham-saham representatif dan diperoleh vektor bobot  $w$  berukuran  $n \times 1$  dengan bobot saham selain saham representatif bernilai nol.

4) *Skenario 4:* Gambar 3d mengilustrasikan skenario 4, dimana dalam tiap kluster, hanya ada tepat satu portofolio sebagai portofolio representatif. Mirip seperti Skenario 3, hanya saja dalam seleksi saham sebagai saham representatif tidak menggunakan nilai *threshold*, melainkan hanya dipilih satu aset dengan bobot tertinggi dari masing-masing kluster untuk menjadi saham representatif. Pada akhirnya, dilakukan optimasi bobot terhadap saham-saham representatif tersebut dan diperoleh vektor bobot  $w$  berukuran  $n \times 1$  dengan bobot saham selain saham representatif bernilai nol.

#### IV. HASIL DAN DISKUSI

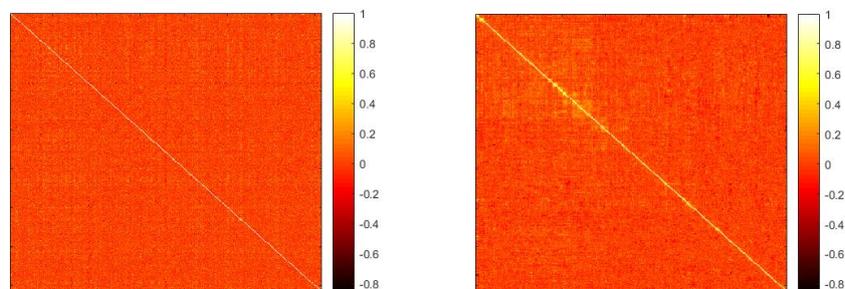
Setelah berhasil mendefinisikan jumlah kluster optimal yaitu  $K = 32$ , perlu dilihat kualitas kluster yang dihasilkan yang dapat ditunjukkan oleh Gambar 4.



Gambar 4: *Silhouette Coefficient*,  $K = 32$

Gambar tersebut memberitahu kita bahwa klusterisasi telah dilakukan dengan baik, dimana  $SC_i$  didominasi oleh nilai positif dan  $SC_i$  dengan nilai negatif terkecil tidak melebihi  $-0.2$ . Sebagai informasi tambahan, rata-rata  $SC$  bernilai  $0.0522$ .

Dampak dari klusterisasi umumnya terlihat saat dilakukan perbandingan matriks korelasi sebelum dan sesudah klusterisasi seperti pada Gambar 5.



(a) Matriks Korelasi Sebelum Klusterisasi      (b) Matriks Korelasi Setelah Klusterisasi

Gambar 5: Matriks Korelasi

Dari Gambar 5a terlihat bahwa setiap korelasi antar saham belum diurutkan dan masih tersebar, sedangkan dari Gambar 5b terlihat bahwa setiap korelasi antar saham telah diurutkan berdasarkan kluster dan saham-saham dengan korelasi tinggi terkumpul pada daerah sudut kiri atas dan pada diagonal matriksnya yang terlihat lebih tegas, yang menandakan kluster yang terbentuk berhasil mengumpulkan saham-saham dengan korelasi tinggi ke dalam kluster yang sama.

Proses yang dilakukan selanjutnya adalah proses konstruksi portofolio. Dalam penelitian ini, dikonstruksi tujuh portofolio yang terdiri dari masing-masing satu portofolio oleh skenario 1, skenario 2, dan skenario 4, dan empat portofolio oleh skenario 3 dengan empat nilai  $threshold = \{0.1, 0.2, 0.3, 0.4\}$ .

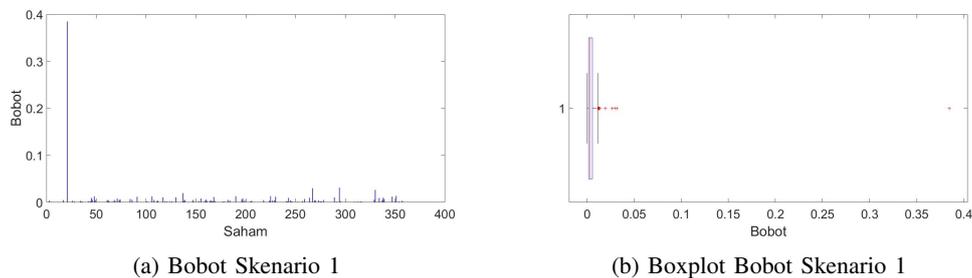
Setelah dilakukan analisis performansi terhadap ketujuh portofolio, nilai harapan  $return E(r_p)$ , variansi  $\sigma_p^2$ , prediksi rasio Sharpe  $S_p$  berdasarkan data *in-sample*, dan rasio Sharpe realisasi  $S'_p$  berdasarkan data *out-sample* portofolio dirangkum dalam Tabel I berikut.

Tabel I: Harapan *Return*, Variansi, dan Rasio Sharpe Portofolio

Skenario	$E(r_p)$	$\sigma_p^2$	$S_p$	$S'_p$
1	0.0010	$1.44 \times 10^{-6}$	0.8695	0.0393
2	0.0021	$10.24 \times 10^{-6}$	0.6369	0.1215
3, <i>threshold</i> = 0.2	0.0021	$10.89 \times 10^{-6}$	0.6513	0.1029
3, <i>threshold</i> = 0.3	0.0021	$12.25 \times 10^{-6}$	0.6135	0.0963
3, <i>threshold</i> = 0.4	0.0021	$14.44 \times 10^{-6}$	0.5603	0.0763
3, <i>threshold</i> = 0.1	0.0022	$10.24 \times 10^{-6}$	0.7041	0.1041
4	0.0022	$13.69 \times 10^{-6}$	0.5821	0.0686

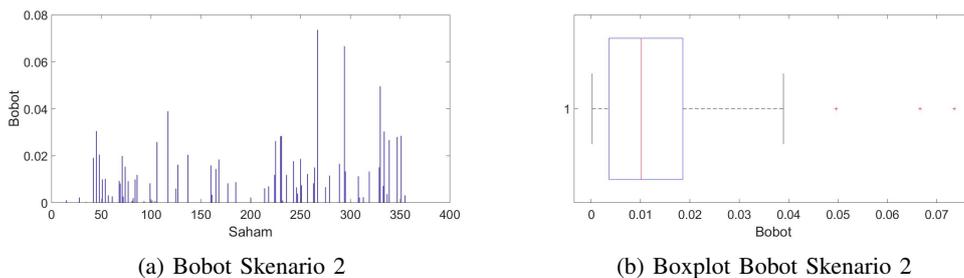
Hasil pada Tabel I menyajikan nilai harapan return, variansi, dan rasio Sharpe  $S_p$  dan  $S'_p$  dari portofolio efisien Markowitz untuk masalah (12), yang dikonstruksikan dengan skenario 1-4. Terdapat tiga tingkat return portofolio untuk dataset diberikan, yaitu  $E(r_p) = 0.1\%$ ,  $0.21\%$ , dan  $0.22\%$ . Portofolio dengan resiko terkecil diberikan oleh skenario-1 dengan tingkat return dan nilai  $S'_p$  terkecil diantara semua skenario. Jelas hasil ini bukan yang menjadi harapan investor untuk portofolio optimal. Yang paling menarik adalah secara berturut-turut pada level harapan return  $0.21\%$  dan  $0.22\%$  skenario-2 dan skenario-3 (*threshold* = 0.1) memberikan nilai risiko terkecil dengan nilai rasio Sharpe  $S_p$  dan  $S'_p$  lebih tinggi diantara portofolio dari semua skenario. Jadi dalam hal ini skenario-2 dan skenario-3 dengan *threshold* bobot 0.1 memberikan hasil diversifikasi portofolio paling optimal diantara skenario lainnya. Terlebih jika dibandingkan dengan skenario-1 yang merupakan skenario diversifikasi portofolio efisien tanpa analisis klaster.

Untuk membantu menganalisis penyebab perbedaan performansi tiga skenario portofolio optimal di tiga tingkat harapan return  $E(r_p)$  yang berbeda, dapat dilihat pada Gambar 6, 7, dan 8 berikut, yang memperlihatkan sebaran nilai bobot portofolio pada setiap skenario yang dilakukan.

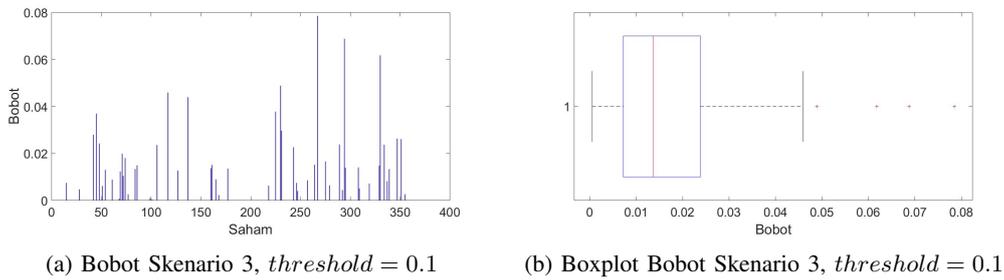


Gambar 6: Bobot Aset Skenario 1

Penyebab dari buruknya performansi portofolio skenario 1 adalah ketergantungannya terhadap saham ke-21 seperti yang digambarkan oleh Gambar 6a dengan bobot sebesar 0.3847 dimana median nilai bobot dalam portofolio ini berdasarkan 6b adalah sebesar 0.0032, sehingga bobot saham ke-21 berjarak sangat jauh dari pusat data.

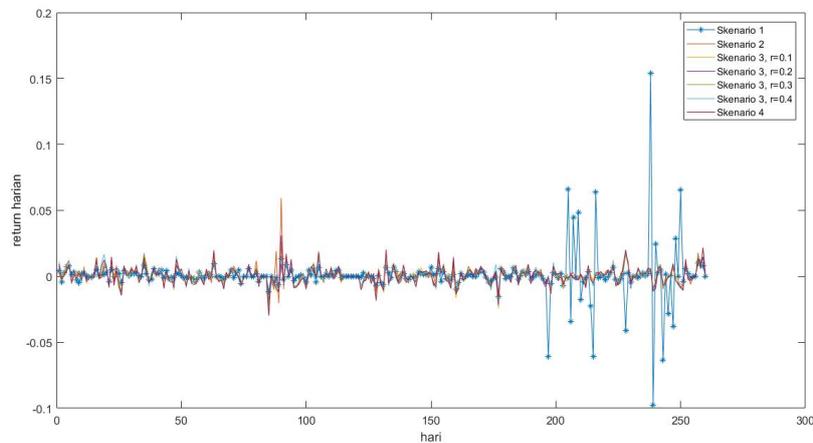


Gambar 7: Bobot Aset Skenario 2



Gambar 8: Bobot aset Skenario 3,  $threshold = 0.1$

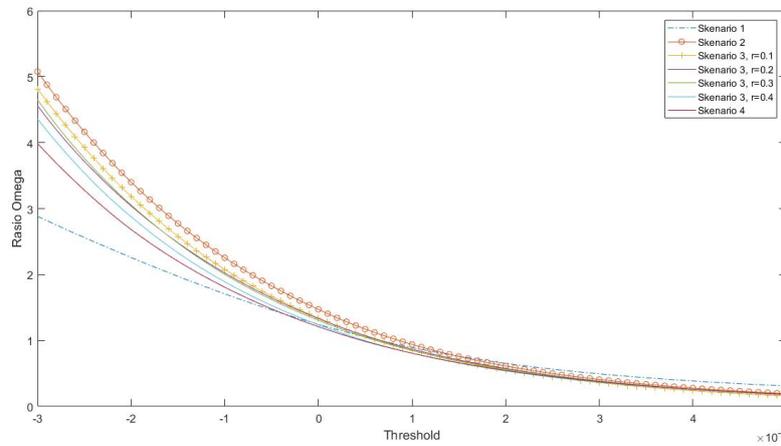
Sedangkan bobot aset dari portofolio skenario 2 dan skenario 3 dengan  $threshold = 0.1$  terdiversifikasi secara lebih merata, seperti yang diperlihatkan oleh Gambar 7a dan 8a. Gambar 7b dan 8b menginformasikan untuk skenario 2 dan skenario 3 dengan  $threshold = 0.1$  hanya terdapat tiga dan empat bobot pencilan dengan pencilan terjauh masing-masingnya sebesar 0.0736 dan 0.0785 dari pusat data yang masing-masing bernilai 0.0101 dan 0.0137. Performansi buruk portofolio dari skenario 1 dapat terlihat pada nilai return harian portofolio, pada Gambar 9. Terlihat bahwa skenario 1 memberikan hasil portofolio paling volatil atau memiliki resiko ketidakpastian return paling tinggi diantara yang lain.



Gambar 9: Return harian portofolio dari setiap skenario.

*Return* portofolio skenario 1 sangat fluktuatif dibandingkan portofolio lainnya pada transaksi di atas hari ke-197. Untuk keenam portofolio lainnya, pergerakan *return*-nya tidak terlalu fluktuatif, sehingga tujuan diversifikasi untuk menurunkan risiko investasi tercapai.

Gambar 10 menunjukkan performansi yang luar biasa dari portofolio-portofolio selain portofolio skenario 1 yang digambarkan dengan garis putus-putus, terutama kedua portofolio unggulan di tiap-tiap kelas  $E(r_p)$  yang memberikan performansi terbaik diantara kelima portofolio saingannya.

Gambar 10: Rasio  $\Omega$ 

Jelas terlihat bahwa portofolio skenario 2, skenario 3 dengan  $r = 0.1$ , skenario 3 dengan  $r = 0.2$ , dan skenario 3 dengan  $r = 0.3$  memiliki kemampuan untuk mencapai bahkan melebihi nilai *threshold*  $r$  yang ditentukan lebih baik dari portofolio skenario 1 untuk  $r \leq 0$ . Bahkan, dengan melihat gambar tersebut, semakin negatif nilai *threshold*-nya, keenam portofolio selain portofolio skenario 1 memberikan manfaat diversifikasi yang luar biasa. Performansi seperti ini yang diharapkan dimiliki oleh suatu portofolio.

## V. KESIMPULAN DAN SARAN

Analisis kluster sudah menjadi tren dalam analisis portofolio karena kemampuannya untuk membantu investor melihat data dari sudut pandang yang baru dan terukur untuk menemukan pola-pola tertentu dalam data yang tidak kasat mata. Analisis kluster yang dilakukan menggunakan  $K = 32$ , menunjukkan kualitas kluster yang baik dalam pengelompokan saham-saham. Hal ini terukur dengan nilai  $SC_i$  yang didominasi oleh nilai positif serta nilai  $SC_i$  negatif terkecil tidak kurang dari  $-0.2$ . Selain itu, hasil baik juga ditunjukkan melalui perbandingan matriks korelasi sebelum dan sesudah klusterisasi.

MVO merupakan rangka kerja unggulan dalam konstruksi dan optimasi portofolio berorientasi *performance-seeking*, namun kehadiran analisis kluster mampu membawa MVO ke tingkat yang lebih tinggi dengan performansi dan analisis yang lebih baik. Pernyataan tersebut dibuktikan oleh penelitian ini, dimana keenam skenario dengan analisis kluster memberikan performansi dan manfaat diversifikasi yang melebihi skenario tanpa analisis kluster. Skenario tanpa analisis kluster, yaitu skenario 1 hanya memberikan  $S'_p$  sebesar 0.0393, yang merupakan nilai  $S'_p$  terendah dibandingkan semua portofolio saingannya. Dalam kelas portofolio dengan  $E(r_p) = 0.0021$ , skenario 2 mengonstruksi portofolio yang paling efisien di kelasnya dengan  $\sigma_p^2 = 10.24 \times 10^{-6}$  dan  $S'_p = 0.1215$ . Sedangkan pada kelas portofolio dengan  $E(r_p) = 0.0022$ , skenario 3 dengan *threshold* = 0.1 menyuguhkan portofolio yang berhasil mengalahkan portofolio saingannya dengan  $\sigma_p^2 = 10.24 \times 10^{-6}$  dan  $S'_p = 0.1041$ . Dari semua portofolio, portofolio hasil konstruksi skenario 2 memberikan performansi terbaik jika dilihat dari rasio  $\Omega$  yang ditandai dengan tingginya nilai rasio  $\Omega$  untuk nilai *threshold* negatif.

Luasnya dunia analisis portofolio investasi, melimpahnya rangka kerja konstruksi dan optimasi portofolio menciptakan luasnya ruang untuk penelitian-penelitian terkait. Penulis menyarankan untuk dilakukannya penelitian lebih lanjut dengan data yang lebih besar, *benchmark* konstruksi dan optimasi portofolio yang lebih kreatif seperti implementasi pengukuran risiko alternatif *semivariance*, *mean-absolute deviation* (MAD) atau *Value at Risk* (VaR), memanfaatkan estimasi galat seperti analisis Bayesian, menerapkan distribusi *stable Paretian* sebagai asumsi distribusi *return*, dan memodelkan dinamika harga aset dengan *time-series* finansial, pohon binomial, atau proses stokastik. Tidak hanya dari kaca mata finansial, penelitian lebih lanjut dapat dilakukan dari segi implementasi metode-metode *data mining*

yang lebih mutakhir, seperti algoritma genetik, *deep learning*, dan pemanfaatan ukuran-ukuran finansial perusahaan sebagai atribut selain korelasi antar saham.

#### PUSTAKA

- [1] T. Calinski and J. Harabasz. A dendrite method for cluster analysis. *Communications in Statistics - Theory and Methods*, 3(1):1–27, 1974.
- [2] Marcos López de Prado. Building diversified portfolios that outperform out of sample. *The Journal of Portfolio Management*, 42(4):59–69, May 2016.
- [3] Jack Clark Francis and Dongcheol Kim. *Modern portfolio theory: foundations, analysis, and new developments website*. Wiley, 2013.
- [4] J. Han, J. Pei, and M. Kamber. *Data Mining: Concepts and Techniques*. The Morgan Kaufmann Series in Data Management Systems. Elsevier Science, 2011.
- [5] Diego León, Arbey Aragón, Javier Sandoval, Germán Hernández, Andrés Arévalo, and Jaime Niño. Clustering algorithms for risk-adjusted portfolio construction. *Procedia Computer Science*, 108:1334–1343, 2017.
- [6] Andrew W. Lo. The statistics of sharpe ratios. *Financial analysts journal*, pages 36–52, 2002.
- [7] R.N. Mantegna. Hierarchical structure in financial markets. *The European Physical Journal B*, 11(1):193–197, September 1999.
- [8] Harry M. Markowitz. Markowitz revisited. *Financial Analysts Journal*, 32(5):47–52, 1976.
- [9] Irma Palupi, Bambang Ari Wahyudi, and Indwiarti Indwiarti. The clustering algorithms approach for decision efficiency in investment portfolio diversification. *7th International Conference on Information and Communication Technology (ICoICT)*, pages 1–6, 2019.
- [10] Bilgehan Tekin and Fatih Burak Gumus. The classification of stocks with basic financial indicators: An application of cluster analysis on the BIST 100 index. *International Journal of Academic Research in Business and Social Sciences*, 7(5), May 2017.
- [11] Joe H. Ward. Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*, 58(301):236–244, March 1963.
- [12] Jin Zhang and Dietmar Maringer. Asset Allocation under Hierarchical Clustering. Working Papers 036, COMISEF, May 2010.