BAB III

METODOLOGI PENELITIAN

3.1 Desain Penelitian

Penelitian ini tergolong sebagai penelitian kuantitatif yang menggabungkan pendekatan deskriptif dan eksploratif. Pendekatan deskriptif bertujuan untuk memberikan gambaran secara sistematis, faktual, dan akurat mengenai karakteristik atau fenomena yang sedang diteliti, sedangkan pendekatan eksploratif digunakan untuk menggali dan memahami lebih dalam terkait aspek-aspek yang belum banyak diketahui sebelumnya mengenai topik penelitian ini (Sugiyono, 2016). Pada kasus penelitian ini, pengelompokan saham berdasarkan indikator indikator fundamental dari saham yaitu imbal hasil pengembalian dan risiko saham yang dilakukan dengan tujuan untuk mendeskripsikan apa yang terjadi pada Emiten anggota Index LQ45 yang merupakan bagian besar dari pasar modal Indonesia yang sangat berpengaruh kepada IHSG.

Penelitian eksploratif di sisi lain merupakan penelitian yang bertujuan untuk mengeksplorasi sebuah fenomena baru dengan metodologi yang baru pula (Riyadhi, 2023). Metodologi baru yang dimaksud pada penelitian ini adalah pengelompokan hibrida dengan metode hirarki *Centroid Linkage* dan metode non-hirarki *K-Means* untuk mengelompokkan saham emiten anggota Index LQ45 di Indonesia selama periode 2023-2024.

3.2 Jenis dan Sumber Data

Jenis data yang digunakan dalam penelitian ini adalah data kuantitatif yang bersumber dari data sekunder. Data sekunder diperoleh secara tidak langsung melalui media atau pihak lain, seperti laporan, publikasi resmi, atau dokumentasi yang telah tersedia. Metode pengumpulan data dilakukan melalui studi pustaka (*literature review*), yaitu dengan menelaah teori-teori yang relevan terkait analisis *K-Means Clustering*, yang diperoleh dari berbagai referensi seperti buku, jurnal

Evonia R. Hutajulu, 2025
IMPLEMENTASI PEMILIHAN PORTOFOLIO OPTIMAL MENGGUNAKAN METODE KLASTERISASI K-MEANS DAN MEAN VARIANCE EFFICIENT PORTOFOLIO (MVEP) (STUDI KASUS INDEKS SAHAM LQ45)
Universitas Pendidikan Indonesia | repository.upi.edu | perpustakaan.upi.edu

ilmiah, dan sumber akademik lainnya (Sugiyono, 2016). Penelitian ini menggunakan data sekunder berupa data harga penutupan saham pada Indeks Saham LQ45 yang diunduh dari website Yahoo Finance, https://www.finance.yahoo.com. Adapun periode harga penutupan saham yang digunakan yakni periode 2023-2024.

Dalam penelitian ini, perangkat lunak yang digunakan adalah Rstudio dengan *package* meliputi:

1. Paket analisis data: Penggunaan paket ggplot2, tidyr, reshape2 pada Rstudio untuk manipulasi data dan penghitungan statistik
2. Paket *clustering:* Penggunaan paket factoextra, Rtools, dplyr, dan clusterSim pada Rstudio untuk mengimplementasikan metode *Centroid Linkage* dan *K-Means Clustering.*
3. Paket Optimisasi: Penggunaan paket quantmod, PerformanceAnalytics, dan quadprog untuk mengimplementasikan model MVEP dengan kendala yang digunakan.

3.3 Variabel Penelitian

Fundamental saham yang digunakan sebagai variabel input penelitian dalam tahap klasterisasi saham adalah imbal hasil saham dan risiko saham. Sedangkan variabel teknikal yang digunakan dalam membangun portofolio dari hasil klaster adalah *Sharpe ratio* seperti yang ditunjukkan pada Tabel 3. 1

Tabel 3. 1 Variabel Penelitian

Variabel	Keterangan	Skala
X_p	Kode Saham	Nominal
$X_{1,p}$	Imbal Hasil Saham	Rasio
$X_{2,p}$	Risiko Saham	Rasio
$X_{3,p}$	*Sharpe Ratio*	Rasio

Selanjutnya struktur data pada penelitian ini disajikan dalam Tabel 3.2 di mana peridoe yang digunakan adalah dari 1 Januari 2023 sampai dengan 31 Desember 2024.

Tabel 3. 2 Struktur Data Penelitian

X_p	$X_{1,p}$	$X_{2,p}$	$X_{3,p}$
X_1	$X_{1,1}$	$X_{2,1}$	$X_{3,1}$
X_2	$X_{1,2}$	$X_{2,2}$	$X_{3,2}$
X_3	$X_{1,3}$	$X_{2,3}$	$X_{3,3}$
...	...	...	...
X_{35}	$X_{1,35}$	$X_{2,35}$	$X_{3,35}$
X_{36}	$X_{1,36}$	$X_{2,36}$	$X_{3,36}$

3.4 Prapemrosan Data

Prapemrosesan data mencakup serangkaian teknik yang diterapkan pada data mentah agar sesuai untuk analisis atau pemodelan (Muslikh dkk., 2023). Proses ini mencakup pembersihan, transformasi, penataan data untuk meningkatkan kualitas, relevansi, dan efisiensi untuk tugas-tugas analisis selanjutnya (Pacovský, 2017). Tujuan utamanya adalah untuk memastikan data bersih dan konsisten, yang sangat penting untuk analisis yang akurat, mengurangi waktu dan sumber daya komputasi yang dibutuhkan untuk *data mining*, dan meningkatkan keandalan dan efisiensi pelatihan model dan hasil.

1) Perhitungan Imbal hasil dan Imbal Hasil Harapan Portofolio

Perhitungan portofolio optimal diawali dengan mengestimasi imbal hasil (return) dari masing-masing saham yang menjadi komponen portofolio. Imbal hasil merujuk pada keuntungan atau pendapatan yang diperoleh dari suatu investasi dalam kurun waktu tertentu, dan digunakan sebagai indikator kinerja aset dalam portofolio tersebut (Jogiyanto, 2010). Misalkan P_{it}, adalah harga aset ke-*i* pada waktu *t* dan diasumsikan bahwa tidak ada pembagian

laba (deviden). Imbal hasil aset ke-i untuk satu periode, dari waktu $t-1$ hingga waktu t adalah (Gubu dkk., 2023):

$$R_{it} = \frac{P_{it} - P_{i(t-1)}}{P_{i(t-1)}} = \frac{P_{it}}{P_{i(t-1)}} - 1, i = 1,2, \dots, n \quad (3.1)$$

dengan mengasumsikan bahwa setiap kemungkinan hasil memiliki peluang yang sama untuk terjadi, maka nilai imbal hasil harapan (*expected return*) dapat dihitung menggunakan rata-rata aritmatika dari seluruh kemungkinan imbal hasil. Secara matematis, imbal hasil harapan dirumuskan sebagai:

$$E[R_{it}] = \frac{1}{T}\sum_{t=1}^{T} r_{it}, i = 1,2, \dots, n \quad (3.2)$$

dengan n adalah periode waktu.

2). Risiko dalam investasi diukur dengan menghitung deviasi standar atau variansi dari imbal hasil, yang merepresentasikan besarnya penyimpangan antara imbal hasil realisasi (*realized return*) dan imbal hasil harapan (*expected return*). Nilai ini mencerminkan volatilitas atau ketidakpastian dari pengembalian investasi. Secara matematis, risiko dapat dihitung dengan rumus sebagai berikut:

$$\sigma_i^2 = \frac{1}{T}\sum_{t-1}^{T} \left(R_{i,t} - E(R_i)\right)^2 \quad (3.3)$$

3.5 Klasterisasi Hibrida

Klasterisasi hibrida adalah sebuah pendekatan yang mengintegrasikan kekuatan dari metode hirarki dan nonhirarki untuk mencapai hasil yang optimal. Pada setiap langkah, sebuah kelompok dipilih berdasarkan kriteria tertentu untuk pembagian lebih lanjut, yang pada akhirnya membentuk jumlah klaster yang diinginkan (Murugesan & Jun, 2011). Pada penelitian ini akan dilakukan penggabungan dua metode klasterisasi yaitu metode *centroid* linkage untuk klasterisasi hirarki dan metode *K-Means* untuk metode nonhirarki.

3.5.1 Algoritma *Centroid Linkage*

Centroid linkage menghitung rata-rata semua objek dalam sebuah klaster. Jarak antara dua klaster ditentukan dengan mengukur jarak antara *centroid* masing-masing. *Centroid* klaster mewakili nilai rata-rata pengamatan di seluruh variabel dalam klaster tertentu. Dengan menggunakan metode ini, *centroid* dihitung ulang setiap kali sebuah klaster baru dibuat, dan terus berlanjut hingga klaster menjadi stabil (Rahmawati, 2012).

Centroid mewakili rata-rata dari semua elemen dalam sebuah klaster. Setiap kali objek baru ditambahkan ke dalam klaster, *centroid* dihitung ulang, sehingga menyebabkan diperbaharui setiap kali ada penambahan. Keunggulan dari metode ini adalah tingkat sensitifitas yang rendah terhadap *outlier* dibandingkan dengan beberapa metode yang lain (Silvi, 2018). Langkah-langkah pada *Cenroid Linkage* sebagai berikut:

1. Perlakukan setiap pengamatan atau indivisu sebagai klasternya sendiri. Buatlah matriks jarak $D = \{d_{ik}\}$ yang merepresentasikan jarak antar klaster.
2. Tentukan jarak terkecil antara dua klaster, yang dilambangkan sebagai d_{uv} (jarak antara kalster u dan v).
3. Gabungkan klaster u dan v, dan perbaharui matriks jarak yang sesuai.
4. Lakukan langkah 2 dan 3 secara berulang hingga $N - 1$ penggabungan selesai, catat jarak pada setiap penggabungan.
5. Pilih nilai *cut off* untuk menentukan jumlah klaster. Gunakan dendogram untuk memvisualisasikan proses dan memutuskan di mana titik potongnya.
6. Tetapkan nama untuk klaster dengan membuat profil karakteristik masing-masing klaster, menganalisis fitur rata-rata anggota dalam setiap klaster.

3.5.2 Algoritma *K-Means Clustering* dan Optimisasi Portofolio Model *Mean-Variance Efficient Portofolio* (MVEP)

1). Standarisasi Data

Standarisasi data diperlukan apabila terdapat perbedaan skala atau satuan antar variabel, terutama ketika variabel memiliki nilai dengan skala yang besar dan berbeda satu sama lain. Salah satu metode yang umum digunakan untuk menstandarisasi data adalah dengan menggunakan Z-score standardization, yaitu mengonversi nilai asli menjadi nilai standar berdasarkan distribusi normal. Nilai Z menunjukkan seberapa jauh sebuah data berada dari rata-rata dalam satuan standar deviasi. (Ghozali, 2018):

$$z = \frac{x_i - \bar{x}}{s} \tag{3.4}$$

dengan:

x_i: data ke-i

$\bar{x}$: rata-rata data

s: simpangan baku

2) Pemilihan k Jumlah Klaster

Metode *Elbow* merupakan teknik yang digunakan untuk menentukan jumlah klaster optimal dalam analisis klaster, khususnya K-Means. Pendekatan ini dilakukan dengan membandingkan jumlah klaster k terhadap nilai *Within-Cluster Sum of Squares* (WCSS), yaitu total jarak kuadrat antara titik-titik data dan *centroid* klasternya. Grafik Elbow menunjukkan penurunan nilai WCSS seiring bertambahnya jumlah klaster. Titik "siku" (*elbow*) pada grafik mengindikasikan jumlah klaster optimal, karena setelah titik tersebut, penurunan WCSS tidak lagi signifikan (Kodinariya & Makwana, 2013)

$$WCSS = \sum_{k=1}^{k} \sum_{X_i \in S_k} (X_i - C_k)^2 \tag{3.5}$$

dengan:

X_i: data ke-i

C_k: *centroid* klaster ke-i

3) Menentukan *centroid* Awal

Secara umum, pemilihan pusat *cluster* ditentukan secara acak di awal (Syakur dkk., 2018). Namun karena pendekatan ini menggabungkan klasterisasi hirarki dan *K-Means*, pusat klaster pada algoritma *K-Means* ditentukan sebagai rata-rata data dalam satu klaster hasil klasterisasi hirarki. Hal ini memastikan pusat klaster dapat langsung ditentukan berdasarkan data yang sesuai dengan jumlah klaster yang diinginkan (BJ, 2015).

4) Perhitungan Jarak Euclidean

Pemilihan awal pusat klaster (*centroid*) dalam metode *K-Means* dilakukan secara acak. Selanjutnya, dihitung jarak antara setiap data terhadap pusat klaster awal tersebut. Jarak antar data, dalam hal ini antar saham, dihitung menggunakan metrik *Euclidean Distance* sebelum proses klasterisasi dimulai. Proses ini bertujuan untuk mengelompokkan data ke dalam klaster berdasarkan kedekatan jaraknya terhadap *centroid* masing-masing (Pratama, 2024):

$$d(\boldsymbol{x}, \boldsymbol{y}) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}\,; i = 1, 2, \dots, n \tag{3.7}$$

5) Kelompokkan seluruh data atau objek ke dalam klaster terdekat. Kedekatan antar objek dihitung berdasarkan jarak antara objek tersebut, dalam kasus ini dengan metode jarak Euclidean. Begitu pula kedekatan suatu data dengan klaster ditentukan dari jarak anatara data tersebut dengan pusat klaster.

6) Perbaharui pusat klaster berdasarkan keanggotaan klaster baru. Pusat klaster dihitung sebagai rata-rata dari semua data atau objek dalam klaster tersebut dengan persamaan:

$$\boldsymbol{v} = \frac{1}{n} \sum_{i=1}^{n} x_i \tag{3.8}$$

dengan:

x_i: objek ke-i

n: jumlah objek anggota klaster

Lakukan kembali langkah ke-4 sampai ke-6 jika posisi *centroid* masih berpindah hingga *centroid* stabil, perulangan dihentikan apabila tidak ada lagi data yang berpindah.

3.6 Perhitungan Bobot Saham Menggunakan Metode MVEP

Dalam penelitian ini, digunakan metode *Mean Variance Efficient Portofolio* (MVEP) sebagai salah satu metode untuk membentuk portofolio yang optimal (Fabozzi, 1999). MVEP dapat dipahami sebagai portofolio dengan varian minimum diantara keseluruh portofolio yang dapat dibentuk. Hal tersebut sama dengan mengoptimalkan bobot $\boldsymbol{w} = [\boldsymbol{w}_1 \dots \boldsymbol{w}_2]^T$ berdasarkan maksimum *mean return* dari varian yang diberikan (Maruddani & Purbowati, 2009).

Permasalahan optimalisasi ini dapat diselesaikan dengan fungsi Lagrange sebagai berikut:

$$L = \boldsymbol{w}^T \sum \boldsymbol{w} + l_1\left(\boldsymbol{\mu}_p - \boldsymbol{w}^T \boldsymbol{\mu}\right) + l_2\left(1 - \boldsymbol{w}^T \mathbf{1}_p\right) \quad 3.9$$

dengan

L : fungsi Lagrange

l : konstanta (pengali) Lagrange

Untuk kasus portofolio dengan varian minimum, tidak ada pembatasan pada *mean* portofolio $(l_1 = 0)$, sehingga pembobotan pada *Mean Variance Efficient Portofolio* adalah:

$$\boldsymbol{w} = \frac{\left(\sum^{-1} \mathbf{1}_p\right)}{\mathbf{1}_N^T \sum^{-1} \mathbf{1}_p} \quad 3.10$$

dengan

$\sum^{-1}$: invers matrik varian-kovarian

$\mathbf{1}_p$: vektor satu dengan dimensi $p \; x \; 1$

3.7 Evaluasi Kinerja Portofolio

Evaluasi terhadap kinerja portofolio merupakan langkah penting dalam manajemen investasi. Salah satu indikator yang umum digunakan untuk mengukur kinerja portofolio adalah rasio Sharpe. Rasio ini menilai seberapa besar kelebihan imbal hasil portofolio dibandingkan dengan tingkat imbal hasil bebas risiko, terhadap total risiko portofolio yang diukur dengan deviasi standar. Semakin tinggi nilai rasio Sharpe, maka semakin baik kinerja portofolio tersebut karena menunjukkan bahwa portofolio mampu menghasilkan imbal hasil yang lebih tinggi untuk setiap unit risiko yang diambil (Sharpe, 1966; Gubu dkk., 2023). *Sharpe ratio* dapat dihitung dengan rumus (Francis, 2013)

$$S_p = \frac{\bar{r}_p - r_f}{\sigma_p} \tag{3.11}$$

dengan

p: portofolio

$\bar{r}_p$: imbal hasil harapan portofolio

r_f: imbal hasil yang tersedia dari sekuritas bebas risiko

σ_p: standar deviasi pengamatan

3.8 Tahap-tahap Analisis Data

Tahapan yang dilakukan dalam menganalisis data pada penelitian ini adalah sebagai berikut:

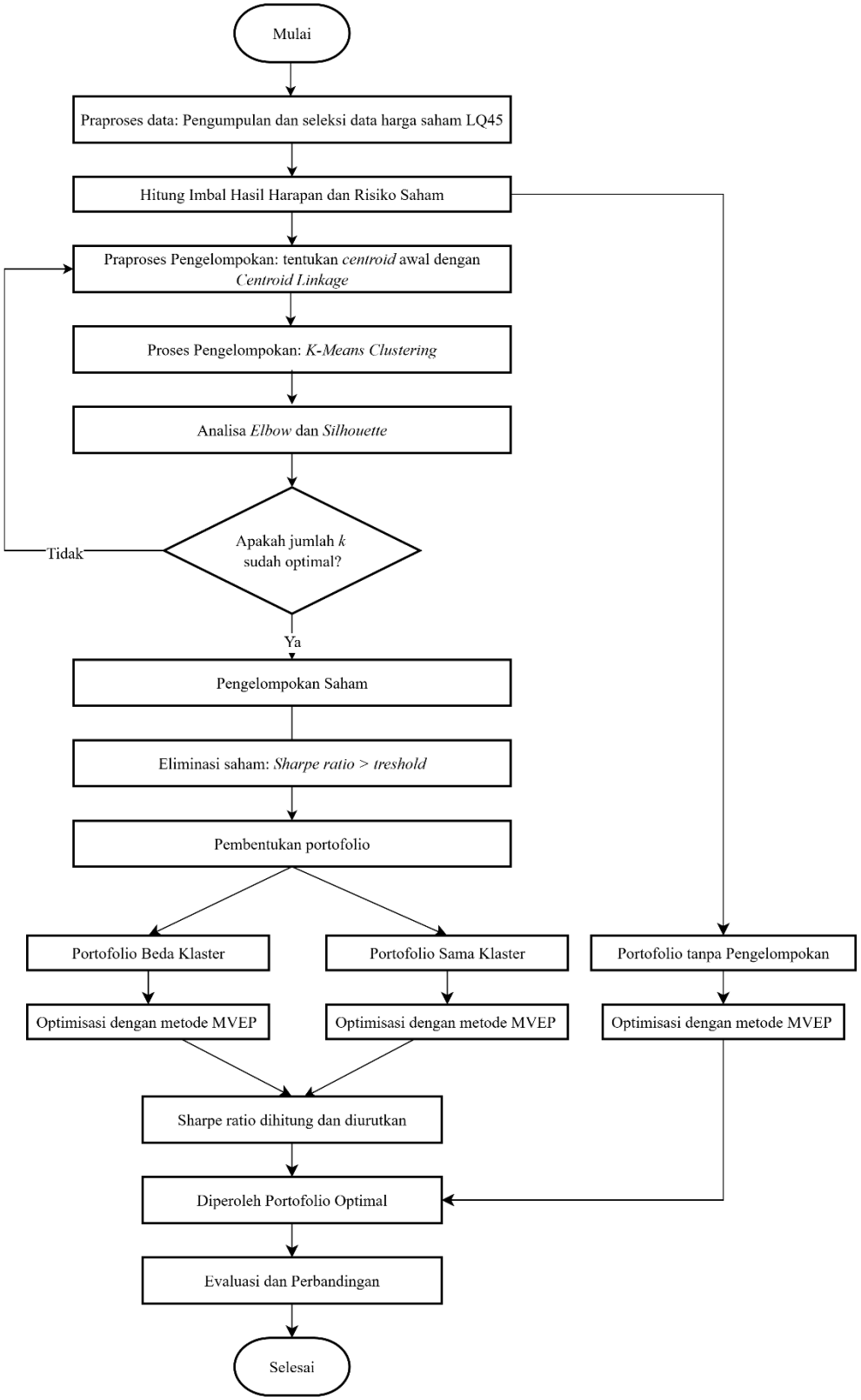
1. Prapores data: Mengumpulkan dan menyeleksi data saham LQ45 dari sumber yang relevan yaitu *finance.yahoo.com*, menghitung imbal hasil dan risiko mingguan selama 2 tahun, serta menghitung *sharpe ratio* selama periode penelitian dari masing-masing emiten Perusahaan anggota Indeks LQ45.
2. Praproses pengelompokan: Melakukan analisis klaster hierarki *Centroid Linkage* untuk menentukan *centroid* awal.
3. Proses pengelompokan: Melakukan pengelompokan saham dengan algoritma *K-Means*

4. Melakukan analisa *elbow method* untuk menentukan nilai *k* optimal dalam proses klasterisasi dengan validasi menggunakan nilai K optimal dalam proses klasterisasi dengan validasi menggunakan *silhouette plot analysis.*
5. Pengelompokan: Melakukan analisa masing-masing klaster dari hasil pengelompokan *K-Means* dan membentuk portofolio dari hasil *cluster.*
6. Eliminasi saham: Menghitung nilai *Sharpe ratio* dari masing-masing portofolio. Rata-rata *sharpe ratio* setiap *cluster* digunakan sebagai *treshold* (ambang batas). Saham dengan *sharpe ratio* di bawah ambang batas dalam setiap klaster akan dieliminasi.
7. Pembentukan portofolio dengan klasterisasi*:* Portofolio dengan klaster berbeda dibuat dengan mengambil satu saham perwakilan dari *k* klaster yang ada, sementara portofolio dengan klaster yang sama dibuat dari sejumlah *k* saham dari satu klaster yang sama.
8. Menentukan portofolio optimal model *MVEP*
    i. Menghitung imbal hasil harapan dan standar deviasi (risiko) saham masing-masing klaster.
    ii. Menghitung kovarians kemudian membuat matriks varians-kovarians.
    iii. Menghitung bobot saham menggunakan metode *Mean Variance Efficient Portofolio.*
    iv. Menghitung imbal hasil harapan portofolio dan varians (Risiko) portofolio masing-masing klaster.
9. Optimisasi portofolio dengan metode MVEP: Bobot saham tiap portofolio dengan *clustering* dioptimalkan menggunakan metode *Mean Variance Efficient Portofolio* (MVEP). Kemudian nilai *sharpe ratio* tiap portofolio dihitung, diurutkan dari yang terbesar dan diambil 3 portofolio terbaik untuk kedua jenis portofolio.

10. Pembentukan dan optimisasi portofolio tanpa *clustering*: Setiap saham dianggap sebagai satu portofolio dan bobot saham-saham dioptimalkan menggunakan MVEP.
11. Evaluasi: Membandingkan performa portofolio yang diisi oleh saham-saham dari kelompok yang berbeda dengan portofolio yang diisi oleh saham-saham dari kelompok yang sama dan portofolio yang dibuat tanpa pengelompokan.
12. Kesimpulan

3.9 Alur Penelitian

Berikut adalah *Flowchart* penelitian dalam tugas akhir ini.


Mulai
Praproses data: Pengumpulan dan seleksi data harga saham LQ45
Hitung Imbal Hasil Harapan dan Risiko Saham
Praproses Pengelompokan: tentukan *centroid* awal dengan *Centroid Linkage*
Proses Pengelompokan: *K-Means Clustering*
Analisa *Elbow* dan *Silhouette*
Apakah jumlah *k* sudah optimal?
Tidak
Ya
Pengelompokan Saham
Eliminasi saham: *Sharpe ratio > treshold*
Pembentukan portofolio
Portofolio Beda Klaster
Portofolio Sama Klaster
Portofolio tanpa Pengelompokan
Optimisasi dengan metode MVEP
Optimisasi dengan metode MVEP
Optimisasi dengan metode MVEP
Sharpe ratio dihitung dan diurutkan
Diperoleh Portofolio Optimal
Evaluasi dan Perbandingan
Selesai

Gambar 3. 1 *Flowchart* Penelitian