

SKRIPSI

**OPTIMASI K-MEANS DAN MARKOWITZ UNTUK
MENINGKATKAN PORTOFOLIO SAHAM PADA
BURSA EFEK INDONESIA**

DISUSUN OLEH

MUTIARA AKBAR NASUTION

2109020113



UMSU

Unggul | Cerdas | Terpercaya

**PROGRAM STUDI TEKNOLOGI INFORMASI
FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI
UNIVERSITAS MUHAMMADIYAH SUMATERA UTARA
MEDAN**

2025

**OPTIMASI K-MEANS DAN MARKOWITZ UNTUK
MENINGKATKAN PORTOFOLIO SAHAM PADA
BURSA EFEK INDONESIA**

SKRIPSI

**Diajukan sebagai salah satu syarat untuk memperoleh gelar Sarjana
Komputer (S.Kom) dalam Program Studi Teknologi Informasi pada Fakultas
Ilmu Komputer dan Teknologi Informasi, Universitas Muhammadiyah
Sumatera Utara**

**Mutiara Akbar Nasution
NPM. 2109020113**

**PROGRAM STUDI TEKNOLOGI INFORMASI FAKULTAS ILMU
KOMPUTER DAN TEKNOLOGI INFORMASI UNIVERSITAS
MUHAMMADIYAH SUMATERA UTARA MEDAN**

2025

LEMBAR PENGESAHAN

Judul Skripsi : OPTIMASI K-MEANS DAN MARKOWITZ
UNTUK MENINGKATKAN PORTOFOLIO
SAHAM PADA BURSA EFEK INDONESIA

Nama Mahasiswa : Mutiara Akbar Nasution

NPM : 2109020113

Program Studi : Teknologi Informasi

Menyetujui
Komisi Pembimbing



(Dr. Al-Khowarizmi, S.Kom., M.Kom)
NIDN. 0127099201

Ketua Program Studi



(Fatma Sari Hutagalung, S.Kom., M.Kom)
NIDN. 0117019301

Dekan



(Dr. Al-Khowarizmi, S.Kom., M.Kom)
NIDN. 0127099201

PERNYATAAN ORISINALITAS

OPTIMASI K-MEANS DAN MARKOWITZ UNTUK MENINGKATKAN PORTOFOLIO SAHAM PADA BURSA EFEK INDONESIA

SKRIPSI

Saya menyatakan bahwa karya tulis ini adalah hasil karya sendiri, kecuali beberapa kutipan dan ringkasan yang masing-masing disebutkan sumbernya.

Medan, 26 April 2025

Yang membuat pernyataan



Mutiara Akbar Nasution
NPM. 2109020113

**PERNYATAAN PERSETUJUAN PUBLIKASI KARYA ILMIAH UNTUK
KEPENTINGAN AKADEMIS**

Sebagai civitas akademika Universitas Muhammadiyah Sumatera Utara, saya bertanda tangan dibawah ini:

Nama : Mutiara Akbar Nasution
NPM : 2109020113
Program Studi : Teknologi Informasi
Karya Ilmiah : Skripsi

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Muhammadiyah Sumatera Utara Hak Bebas Royalti Non-Eksekutif (*Non-Exclusive Royalty free Right*) atas penelitian skripsi saya yang berjudul:

OPTIMASI K-MEANS DAN MARKOWITZ UNTUK MENINGKATKAN PORTOFOLIO SAHAM PADA BURSA EFEK INDONESIA

Beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti Non-Eksekutif ini, Universitas Muhammadiyah Sumatera Utara berhak menyimpan, mengalih media, memformat, mengelola dalam bentuk database, merawat dan mempublikasikan Skripsi saya ini tanpa meminta izin dari saya selama tetap mencantumkan nama saya sebagai penulis dan sebagai pemegang dan atau sebagai pemilik hak cipta.

Demikian pernyataan ini dibuat dengan sebenarnya.

Medan, April 2025

Yang membuat pernyataan



Mutiara Akbar Nasution
NPM. 2109020113

RIWAYAT HIDUP

DATA PRIBADI

Nama Lengkap : Mutiara Akbar Nasution
Tempat dan Tanggal Lahir : Medan, 03 Maret 2004
Alamat Rumah : Jl. Bukit Barisan, Gg. Kelabu No. 7
Telepon/Fax/HP : 082161800054
E-mail : mutiaraakbarnst03@gmail.com
Instansi Tempat Kerja : -
Alamat Kantor : -

DATA PENDIDIKAN

SD	: SD Muhammadiyah 18 Medan	TAMAT: 2015
SMP	: SMP Muhammadiyah 01 Medan	TAMAT: 2018
SMA	: SMA Muhammadiyah 01 Medan	TAMAT: 2021
S1	: Universitas Negeri Medan (Bisnis Digital)	TAMAT: 2025
S1	: Universitas Muhammadiyah Sumatera Utara (TI)	TAMAT: 2025

KATA PENGANTAR



Puji syukur ke hadirat Allah SWT yang telah melimpahkan rahmat dan karunia-Nya sehingga penulis dapat menyelesaikan skripsi dengan judul “*Optimasi K-Means Dan Markowitz Untuk Meningkatkan Portofolio Saham Pada Bursa Efek Indonesia*”. Laporan skripsi ini disusun untuk memenuhi salah satu syarat dalam memperoleh gelar Sarjana Komputer (S.Kom) pada Program Studi Teknologi Informasi, Fakultas Ilmu Komputer dan Teknologi Informasi, Universitas Muhammadiyah Sumatera Utara.

Dengan tulus, Penulis mengucapkan terima kasih kepada kedua orang tua tercinta, **Prof. Dr. H. Elfrianto Nst., S.Pd., M.Pd.** dan **Hj. Subhidawarni Pasaribu S.Ag.**, atas segala doa, dukungan moral dan materi, serta semangat yang tak pernah henti diberikan kepada Penulis. Tanpa mereka, penyusunan skripsi ini tidak akan berjalan sebagaimana mestinya. Penulis juga menyadari bahwa penyusunan skripsi ini tidak terlepas dari bantuan, bimbingan, serta dukungan dari berbagai pihak. Oleh karena itu, dengan segala kerendahan hati, Penulis menyampaikan terima kasih dan penghargaan yang sebesar-besarnya kepada:

1. Bapak Prof. Dr. Agussani, MAP, selaku Rektor Universitas Muhammadiyah Sumatera Utara
2. Bapak Prof. Dr. Muhammad Arifin, S.H., M.Hum, Bapak Prof. Dr. Akrim, M.Pd, dan Bapak Assoc. Prof. Dr. Rudianto, S.Sos., M.Si, selaku Wakil Rektor I, II, dan III Universitas Muhammadiyah Sumatera Utara

3. Bapak Dr. Al-Khowarizmi, S.Kom., M.Kom., selaku Dekan Fakultas Ilmu Komputer dan Teknologi Informasi dan juga Dosen Pembimbing Penulis
4. Bapak Halim Maulana, S.T., M.Kom., selaku Wakil Dekan I Fakultas Ilmu Komputer dan Teknologi Informasi
5. Bapak Dr. Lutfi Basit, S.Sos., M.I.Kom., selaku Wakil Dekan III Fakultas Ilmu Komputer dan Teknologi Informasi
6. Ibu Fatma Sari Hutagalung, S.Kom., M.Kom., selaku Ketua Program Studi Teknologi Informasi
7. Bapak Mhd Basri, S.Si, M.Kom., selaku Sekretaris Program Studi Teknologi Informasi
8. Seluruh Civitas Akademik Universitas Muhammadiyah Sumatera Utara
9. Teman-teman seperjuangan di Program Studi Teknologi Informasi yang telah menjadi tempat berbagi ilmu, semangat, dan tawa selama proses perkuliahan hingga penyusunan skripsi ini

Penulis menyadari bahwa skripsi ini masih jauh dari sempurna. Oleh karena itu, segala kritik dan saran yang membangun sangat Penulis harapkan demi perbaikan di masa mendatang. Semoga karya ini dapat memberikan manfaat, khususnya dalam bidang teknologi informasi dan pengembangan ilmu pengetahuan.

Medan, 20 April 2025
Penulis

Mutiara Akbar Nasution

ABSTRAK

Investasi di pasar saham menuntut pendekatan analisis berbasis data untuk membentuk portofolio yang optimal. Penelitian ini bertujuan untuk mengoptimalkan portofolio saham di Bursa Efek Indonesia dengan mengombinasikan metode K-Means Clustering dan Model Markowitz. Data yang digunakan mencakup harga penutupan saham (Close), rasio Price-to-Earnings (P/E), Price-to-Book Value (PBV), dan Return on Equity (ROE) pada periode 2019 hingga 2024. Proses penelitian dimulai dengan tahap preprocessing data, yaitu normalisasi, perhitungan return, dan volatilitas saham. Selanjutnya, penerapan K-Means Clustering dilakukan untuk mengelompokkan saham berdasarkan kesamaan karakteristik fundamental dan teknikal. Penentuan jumlah kluster menggunakan metode Elbow Method. Setelah kluster terbentuk, dilakukan seleksi satu saham terbaik dari masing-masing kluster berdasarkan rata-rata return tahunan tertinggi. Saham-saham terpilih kemudian dioptimasi menggunakan Model Markowitz dengan pendekatan Mean-Variance Optimization untuk menentukan bobot investasi optimal. Evaluasi portofolio dilakukan melalui simulasi backtest menggunakan data tahun 2024 dan dibandingkan dengan performa Indeks LQ45 sebagai benchmark. Hasil penelitian menunjukkan bahwa kombinasi K-Means Clustering dan Model Markowitz menghasilkan portofolio dengan return yang lebih tinggi dan risiko yang lebih terkontrol dibandingkan benchmark. Pendekatan ini juga terbukti meningkatkan efektivitas diversifikasi dan pengelolaan risiko portofolio. Penelitian ini memberikan kontribusi terhadap penerapan teknik data mining dan optimasi kuantitatif dalam manajemen portofolio saham di pasar modal Indonesia, serta memberikan alternatif strategi investasi berbasis teknologi kepada investor dan akademisi.

Kata Kunci: K-Means Clustering, Model Markowitz, Optimasi Portofolio, Data Mining, Bursa Efek Indonesia

ABSTRACT

Investment in the stock market demands a data-driven analytical approach to construct an optimal portfolio. This study aims to optimize stock portfolios on the Indonesia Stock Exchange by combining K-Means Clustering and the Markowitz Model. The data used includes stock closing prices (Close), Price-to-Earnings (P/E) Ratio, Price-to-Book Value (PBV), and Return on Equity (ROE) over the period 2019 to 2024. The research process begins with data preprocessing, including normalization, calculation of returns, and stock volatility. Subsequently, K-Means Clustering is applied to group stocks based on the similarity of their fundamental and technical characteristics. The number of clusters is determined using the Elbow Method. After clustering, the best-performing stock from each cluster is selected based on the highest average annual return. The selected stocks are then optimized using the Markowitz Model with a Mean-Variance Optimization approach to determine the optimal investment weights. Portfolio evaluation is conducted through backtest simulations using 2024 data and compared with the performance of the LQ45 Index as a benchmark. The results show that the combination of K-Means Clustering and the Markowitz Model produces a portfolio with higher returns and better risk control compared to the benchmark. This approach also proves to enhance the effectiveness of portfolio diversification and risk management. This research contributes to the application of data mining techniques and quantitative optimization in stock portfolio management in the Indonesian capital market, offering an alternative technology-based investment strategy for investors and academics.

Keywords: K-Means Clustering, Markowitz Model, Portfolio Optimization, Data Mining, Indonesia Stock Exchange

DAFTAR ISI

KATA PENGANTAR	i
ABSTRAK.....	iii
<i>ABSTRACT</i>	iv
DAFTAR ISI.....	v
DAFTAR TABEL.....	vii
DAFTAR GAMBAR	viii
BAB I PENDAHULUAN	1
1.1. Latar Belakang Masalah	1
1.2. Rumusan Masalah	3
1.3. Batasan Masalah.....	4
1.4. Tujuan Penelitian.....	4
1.5. Manfaat Penelitian.....	4
BAB II LANDASAN TEORI	6
2.1. Investasi dan Portofolio Saham	6
2.1.1. Definisi dan Konsep Dasar Investasi Saham.....	6
2.1.2. Teori Portofolio dan Diversifikasi Risiko	7
2.2. Data Mining dalam Analisis Saham	7
2.2.1. Konsep dan Manfaat Data Mining.....	7
2.2.2. Penerapan Data Mining dalam Dunia Keuangan	9
2.3. K-Means Clustering	10
2.3.1. Konsep dan Prinsip K-Means Clustering.....	10
2.3.2. Langkah-Langkah Algoritma K-Means	13
2.3.3. Penerapan K-Means Clustering dalam Analisis Saham	14
2.3.4. K-Means dan Model Markowitz dalam Optimasi Portofolio	15
2.4. Model Markowitz.....	17
2.4.1. Konsep Dasar Teori Portofolio Markowitz.....	17
2.4.2. Expected Return, Variance, dan Covariance dalam Portofolio	19
2.4.3. Efisien Frontier dan Optimalisasi Risiko-Return	21
2.4.4. Integrasi dengan K-Means Clustering dalam Pemilihan Saham	23
2.5. Python dalam Analisis Data Keuangan	26
2.5.1. Pustaka Python yang Digunakan	27
2.5.2. Implementasi Data Mining dengan Python.....	28
2.6. Penelitian Terdahulu	29
BAB III METODOLOGI PENELITIAN	33
3.1. Jenis Penelitian.....	33
3.2. Sumber Data dan Teknik Pengumpulan Data	34
3.2.1. Sumber Data	34
3.2.2. Teknik Pengumpulan Data	35
3.3. Teknik Analisis Data	36
3.3.1. Preprocessing Data	36
3.3.2. Penerapan K-Means Clustering	36
3.3.3. Pemilihan Saham dari Setiap Klaster.....	39
3.3.4. Optimasi Portofolio dengan Model Markowitz.....	39
3.3.5. Evaluasi Portofolio	41
3.4. Alat dan Perangkat Lunak	42
3.4.1. Google Colab sebagai Lingkungan Pengembangan	42
3.4.2. Pustaka Python yang Digunakan	42
3.4.3. Instalasi dan Import Pustaka di Google Colab	43

3.4.4.	Alur Pemrograman di Google Colab	43
3.5.	Arsitektur Penelitian	44
3.6.	Waktu Penelitian	46
BAB IV HASIL DAN PEMBAHASAN		47
4.1.	Deskripsi Data	47
4.2.	Pengelompokan Saham dengan K-Means Clustering	56
4.2.1.	Normalisasi dan Seleksi Fitur	56
4.2.2.	Penentuan Jumlah Kluster	58
4.2.3.	Hasil Clustering dan Karakteristik Kluster	60
4.3.	Seleksi Saham dari Setiap Kluster	62
4.3.1.	Kriteria Seleksi	62
4.3.2.	Implementasi Seleksi Saham Otomatis	62
4.3.3.	Hasil Seleksi Saham	64
4.4.	Optimasi Portofolio dengan Model Markowitz	65
4.4.1.	Persiapan Data untuk Optimasi	65
4.4.2.	Persiapan Data untuk Optimasi	67
4.4.3.	Hasil Bobot Portofolio	69
4.5.	Simulasi Kinerja Portofolio Tahun 2024	71
4.5.1.	Metode Simulasi Backtest	71
4.5.2.	Hasil dan Interpretasi	74
4.6.	Perbandingan Kinerja Portofolio dengan Indeks LQ45	75
4.6.1.	Metode Perbandingan	75
4.6.2.	Hasil Perbandingan	76
BAB V KESIMPULAN DAN SARAN		78
5.1.	Kesimpulan	78
5.2.	Saran	79
DAFTAR PUSTAKA		81

DAFTAR TABEL

Tabel 2.1. Penelitian Terdahulu	30
Tabel 3.1. Data Fundamental	34
Tabel 3.2. Harga Close Saham	35
Tabel 4.1. Data Gabungan Teknikal dan Fundamental.....	56
Tabel 4.2. Hasil Klasterisasi.....	61
Tabel 4.3. Interpretasi Gaya Investasi Setiap Kluster	61
Tabel 4.4. Hasil Seleksi Saham	64
Tabel 4.5. Bobot Investasi.....	70

DAFTAR GAMBAR

Gambar 3.1. General Architecture	44
Gambar 4.1. Pengambilan Harga Penutupan Saham	48
Gambar 4.2. Perhitungan Return dan Volatilitas Saham	50
Gambar 4.3. Pembuatan Data Fundamental	52
Gambar 4.4. Penggabungan Fitur Teknikal dan Fundamental	54
Gambar 4.5. Preprocessing Data untuk Clustering	57
Gambar 4.6. Elbow Method untuk Menentukan Jumlah Klaster	58
Gambar 4.7. Grafik Elbow Method	59
Gambar 4.8. Kode Program Klastering	60
Gambar 4.9. Kode Otomatisasi Pemilihan Terbaik Setiap Klaster	63
Gambar 4.10. Pengambilan Data dan Perhitungan Return	66
Gambar 4.11. Persiapan Data	67
Gambar 4.12. Program Visualisasi	69
Gambar 4.13. Hasil Optimasi Markowitz	70
Gambar 4.14. Simulasi Backtest Portofolio Tahun 2024	72
Gambar 4.15. Hasil Simulasi	74
Gambar 4.16. Program Perbandingan Kinerja Portofolio dan Indeks LQ45	75
Gambar 4.17. Perbandingan Kinerja	77

BAB I

PENDAHULUAN

1.1. Latar Belakang Masalah

Investasi di pasar saham telah menjadi salah satu strategi utama bagi investor dalam mengalokasikan dana mereka dengan tujuan memperoleh keuntungan yang optimal. Namun, kompleksitas pasar saham yang dipengaruhi oleh berbagai faktor ekonomi, fundamental perusahaan, serta volatilitas harga saham menuntut adanya strategi investasi yang lebih sistematis dan berbasis data (Utami et al., 2019). Oleh karena itu, diperlukan metode yang lebih sistematis untuk membangun portofolio yang optimal dan mampu meminimalkan risiko tanpa mengorbankan potensi keuntungan.

Salah satu pendekatan yang sering digunakan dalam optimasi portofolio adalah teori portofolio yang dikembangkan oleh Markowitz, yang dikenal sebagai Mean-Variance Portfolio Theory (MVP). Model ini bertujuan untuk membentuk portofolio dengan risiko minimal untuk tingkat pengembalian tertentu melalui diversifikasi aset (Pratama et. al., 2024).

Meskipun model Markowitz telah banyak digunakan dalam pengelolaan portofolio, pendekatan ini memiliki beberapa keterbatasan, terutama dalam hal seleksi saham yang optimal. Diversifikasi yang tidak efektif sering kali menyebabkan investor menghadapi risiko yang lebih tinggi dari yang diharapkan (Sukamto et al., 2023). Oleh karena itu, diperlukan pendekatan tambahan yang dapat meningkatkan efektivitas model Markowitz, salah satunya dengan menerapkan teknik clustering dalam pemilihan saham.

Salah satu metode data mining yang dapat diterapkan dalam optimasi portofolio adalah K-Means Clustering. Metode ini telah terbukti efektif dalam mengelompokkan saham berdasarkan karakteristik tertentu seperti return historis, volatilitas, serta rasio keuangan fundamental seperti Price-to-Earnings (P/E) Ratio dan Price-to-Book Value (PBV) (Tohendry & Jollyta, 2023). Dengan mengelompokkan saham ke dalam beberapa klaster yang homogen, investor dapat memilih saham dari tiap klaster untuk meningkatkan diversifikasi dan mengurangi risiko portofolio.

Penelitian sebelumnya menunjukkan bahwa kombinasi K-Means Clustering dan Model Markowitz dapat meningkatkan optimasi portofolio dengan cara menyeleksi saham berdasarkan karakteristik yang lebih homogen. Utami et al. (2019) mengemukakan bahwa dengan menambahkan teknik clustering pada proses seleksi saham, portofolio yang dihasilkan dapat memiliki risiko yang lebih rendah dengan return yang lebih stabil. Selain itu, penelitian yang dilakukan oleh Pratama et al. (2024) juga menunjukkan bahwa Mean Variance Efficient Portfolio (MVEP) yang dikombinasikan dengan K-Means dapat menghasilkan bobot investasi yang lebih optimal dengan risiko lebih terkontrol.

Sejalan dengan perkembangan teknologi dan pemanfaatan Python dalam analisis data keuangan, metode ini dapat diimplementasikan dalam penelitian untuk melakukan analisis berbasis komputasi yang lebih efisien dan akurat. Dengan menggunakan pustaka seperti Scikit-Learn, Pandas, dan PyPortfolioOpt, implementasi K-Means Clustering dapat dilakukan untuk mengelompokkan saham secara otomatis berdasarkan pola yang ditemukan dalam dataset, sementara Model

Markowitz dapat digunakan untuk mengoptimalkan bobot investasi berdasarkan return dan risiko yang diestimasi dari klaster yang terbentuk (Sukanto et. al., 2023).

Penelitian ini bertujuan untuk menerapkan K-Means Clustering dalam mengelompokkan saham berdasarkan karakteristik fundamental dan teknikal, sehingga saham dengan pola yang serupa dapat dikelompokkan ke dalam klaster yang lebih homogen. Dengan pengelompokan ini, proses seleksi saham dalam pembentukan portofolio dapat dilakukan dengan lebih sistematis. Selanjutnya, hasil dari proses clustering akan digunakan dalam Model Markowitz untuk membentuk portofolio optimal yang mempertimbangkan keseimbangan antara return dan risiko investasi.

Kombinasi kedua metode ini diharapkan dapat meningkatkan efektivitas diversifikasi portofolio, sehingga risiko dapat diminimalkan tanpa harus mengorbankan potensi keuntungan. Selain itu, penelitian ini juga akan menganalisis sejauh mana pendekatan ini mampu mengoptimalkan return portofolio dibandingkan dengan strategi investasi konvensional, memberikan wawasan baru bagi investor dalam mengelola portofolio saham secara lebih efisien.

Dengan demikian, penelitian ini diharapkan dapat memberikan kontribusi dalam pengembangan strategi investasi berbasis data mining yang lebih efektif dan dapat diterapkan oleh investor dalam mengambil keputusan investasi di Bursa Efek Indonesia (IDX).

1.2. Rumusan Masalah

Bagaimana implementasi kombinasi K-Means Clustering dan Model Markowitz menggunakan Python dapat dioptimalkan untuk membentuk portofolio saham yang efisien berdasarkan data fundamental dan teknikal?

1.3. Batasan Masalah

Untuk memastikan penelitian ini terfokus dan memiliki hasil yang dapat diinterpretasikan dengan jelas, penelitian ini memiliki beberapa batasan sebagai berikut:

1. Periode waktu pengambilan data: Data saham yang digunakan dalam penelitian ini akan mencakup periode 2019–2024.
2. Jumlah saham yang dianalisis: Hanya mencakup saham yang terdaftar dalam Bursa Efek Indonesia (IDX) dengan likuiditas tinggi seperti indeks LQ45.
3. Faktor eksternal: Penelitian ini tidak mempertimbangkan faktor eksternal seperti sentimen pasar, kebijakan ekonomi, atau peristiwa global.

1.4. Tujuan Penelitian

Tujuan dari penelitian ini adalah untuk mengimplementasikan dan mengevaluasi kombinasi algoritma K-Means Clustering dan Model Markowitz menggunakan Python dalam membentuk portofolio saham yang optimal berdasarkan data fundamental dan teknikal.

1.5. Manfaat Penelitian

1. Bagi Peneliti

Penelitian ini memberikan kontribusi dalam pengembangan analisis investasi berbasis data mining, khususnya dalam penerapan K-Means Clustering untuk seleksi saham dan Model Markowitz untuk optimasi portofolio. Selain itu, penelitian ini juga memberikan pengalaman dalam mengolah data saham secara kuantitatif menggunakan Python, yang dapat digunakan dalam penelitian lebih

lanjut di bidang financial technology (FinTech) dan data science. Hasil penelitian ini dapat menjadi referensi bagi akademisi yang tertarik dalam studi pengelolaan investasi berbasis teknologi dan model kuantitatif.

2. Bagi Investor

Penelitian ini membantu investor dalam menyusun strategi investasi yang lebih terstruktur dengan menggunakan pendekatan berbasis data. Dengan K-Means Clustering, investor dapat mengelompokkan saham berdasarkan karakteristik fundamental dan teknikal, sehingga memudahkan dalam pemilihan saham yang lebih homogen. Sementara itu, Model Markowitz memberikan pendekatan kuantitatif dalam membangun portofolio yang optimal dengan mempertimbangkan keseimbangan antara return dan risiko. Dengan adanya penelitian ini, investor, terutama yang masih pemula, dapat memahami bagaimana kombinasi teknik ini dapat membantu dalam diversifikasi portofolio dan pengelolaan risiko investasi.

3. Bagi Universitas

Penelitian ini berkontribusi dalam memperkaya referensi akademik di bidang data mining, keuangan, dan teknologi informasi. Hasil penelitian ini dapat digunakan oleh mahasiswa dan dosen sebagai bahan ajar atau referensi untuk penelitian lebih lanjut di bidang machine learning dalam keuangan. Selain itu, penelitian ini juga dapat mendukung pengembangan kurikulum di bidang financial technology, terutama dalam implementasi algoritma analisis investasi menggunakan bahasa pemrograman Python. Dengan adanya penelitian ini, universitas dapat mendorong inovasi di bidang teknologi keuangan serta meningkatkan kualitas publikasi ilmiah yang berorientasi pada analisis data dan optimasi portofolio saham.

BAB II

LANDASAN TEORI

2.1. Investasi dan Portofolio Saham

2.1.1. Definisi dan Konsep Dasar Investasi Saham

Investasi saham semakin mudah dilakukan seiring dengan pesatnya penyebaran informasi, yang memungkinkan investor mengambil keputusan lebih akurat. Investasi dapat didefinisikan sebagai komitmen sejumlah dana atau sumber daya lainnya pada saat ini untuk memperoleh manfaat di masa depan (Samsudin et al., 2023).

Saham merupakan tanda penyertaan modal seseorang atau badan usaha dalam sebuah perusahaan atau perseroan terbatas (Wardhani et al., 2022). Pasar modal berfungsi sebagai tempat pertemuan antara investor dan emiten dalam bursa efek, yang menjadi sarana bagi perusahaan publik untuk menawarkan instrumen keuangan jangka panjang (Nuha et al., 2023). Investasi saham adalah kegiatan menanamkan modal di perusahaan yang terdaftar di Bursa Efek Indonesia dengan tujuan memperoleh keuntungan, baik dari kenaikan harga saham maupun dividen. Investasi ini juga dapat menjadi alternatif sebagai instrumen investasi jangka panjang (Jum'an, 2024).

Namun, sebelum berinvestasi, seorang investor perlu melakukan analisis mendalam sebagai bentuk kehati-hatian dalam memilih saham. Risiko yang menyertai investasi dapat diminimalkan dengan melakukan analisis fundamental dan teknikal, memahami mekanisme pasar modal, serta memprediksi tren saham dan pergerakan harga yang selalu berubah (Ady et al., 2022).

2.1.2. Teori Portofolio dan Diversifikasi Risiko

Pada tahun 1952, Harry Markowitz memperkenalkan teori portofolio modern, yang mengubah cara investor dalam menyusun strategi investasi. Teori ini menekankan bahwa investor sebaiknya mempertimbangkan portofolio secara keseluruhan daripada hanya menganalisis setiap aset secara individual (Markowitz, 1952). Dengan model perhitungan yang dikembangkan, investor dapat membentuk portofolio yang memaksimalkan keuntungan meskipun dengan tingkat risiko yang sama. Teori ini menjadi dasar dalam pengelolaan portofolio hingga saat ini.

Salah satu konsep utama dalam teori ini adalah diversifikasi, yaitu strategi untuk mengurangi risiko investasi dengan mengalokasikan dana ke berbagai aset (Simatupang et al., 2024). Diversifikasi didasarkan pada hukum angka besar, yang menyatakan bahwa risiko portofolio dapat dikurangi dengan menambahkan lebih banyak jenis saham yang memiliki korelasi rendah satu sama lain (Samsudin et al., 2023). Dengan cara ini, investor dapat mengelola risiko secara lebih efektif tanpa harus mengorbankan potensi pengembalian investasi.

2.2. Data Mining dalam Analisis Saham

2.2.1. Konsep dan Manfaat Data Mining

Data mining merupakan teknik dalam ilmu komputer yang bertujuan untuk mengekstrak informasi berharga dari kumpulan data yang besar. Dalam analisis keuangan, data mining digunakan untuk mengidentifikasi pola tersembunyi dalam data historis yang dapat dimanfaatkan untuk pengambilan keputusan investasi yang lebih baik (Sukamto et al., 2023).

Dalam dunia investasi saham, data mining memiliki peran penting dalam membantu investor dan trader memahami pola pergerakan harga saham. Dengan

menggunakan teknik analisis yang tepat, data mining dapat digunakan untuk menemukan faktor-faktor yang mempengaruhi pasar dan merancang strategi investasi yang lebih optimal (Ariyatma & Fahmi, 2023). Selain itu, teknologi ini juga dapat membantu dalam penyaringan saham berdasarkan karakteristik tertentu, seperti saham yang memiliki rasio dividen tinggi atau kinerja fundamental yang baik.

Salah satu manfaat utama dari penerapan data mining dalam analisis saham adalah meningkatkan akurasi prediksi harga saham di masa mendatang. Dengan teknik seperti regresi linier dan machine learning, investor dapat memperkirakan pergerakan harga dengan lebih akurat, sehingga mengurangi risiko kerugian akibat ketidakpastian pasar (Inaku & Chandra, 2023). Selain itu, data mining juga dapat digunakan untuk mengelompokkan saham berdasarkan karakteristik fundamentalnya, seperti yang diterapkan dalam metode K-Means, yang memungkinkan investor memilih emiten yang sesuai dengan strategi investasi mereka (Sukamto et al., 2023).

Selain manfaat yang telah disebutkan, data mining juga memungkinkan analisis sentimen pasar berdasarkan berita keuangan dan media sosial. Dengan memahami bagaimana berita dan tren sosial mempengaruhi pasar, investor dapat membuat keputusan investasi yang lebih baik dan mengurangi dampak volatilitas pasar yang tidak terduga (Usman et al., 2024). Kemampuan data mining dalam mengolah berbagai sumber informasi menjadikannya alat yang sangat berguna dalam dunia keuangan modern.

2.2.2. Penerapan Data Mining dalam Dunia Keuangan

Dalam dunia keuangan, penerapan data mining semakin luas dan digunakan dalam berbagai aspek analisis pasar modal. Salah satu metode yang paling umum digunakan dalam prediksi harga saham adalah regresi linier. Metode ini membantu mengidentifikasi hubungan antara variabel-variabel seperti harga pembukaan, harga tertinggi, harga terendah, dan volume perdagangan dengan harga saham di masa mendatang. Studi yang dilakukan oleh Usman et al. (2024) menunjukkan bahwa regresi linier berganda mampu memberikan prediksi harga saham yang cukup akurat, terutama dalam kasus saham Bank Rakyat Indonesia (BRI).

Selain regresi linier, metode Long Short-Term Memory (LSTM) juga semakin banyak diterapkan dalam analisis harga saham. LSTM merupakan salah satu model deep learning yang dapat menangkap pola pergerakan harga saham berdasarkan data historis dalam jangka waktu yang panjang. Penelitian oleh Inaku & Chandra (2023) menunjukkan bahwa LSTM mampu memberikan prediksi harga saham dengan tingkat error yang rendah, sehingga dapat digunakan sebagai alat bantu dalam strategi investasi yang lebih canggih.

Pendekatan lain yang sering digunakan dalam data mining untuk saham adalah K-Means Clustering. Metode ini digunakan untuk mengelompokkan saham berdasarkan karakteristik fundamentalnya, seperti rasio dividen, profit margin, dan return on equity (ROE). Hasil penelitian Sukamto et al. (2023) menunjukkan bahwa pengelompokan saham berdasarkan K-Means dapat membantu investor dalam menentukan pilihan saham yang lebih sesuai dengan profil risiko mereka.

Selain itu, multiple regression juga telah diterapkan dalam analisis saham untuk meningkatkan akurasi prediksi. Studi yang dilakukan oleh Ariyatma & Fahmi

(2023) menggunakan multiple regression untuk memprediksi harga saham Netflix dan menemukan bahwa model ini dapat menghasilkan prediksi yang lebih akurat dengan tingkat kesalahan yang kecil. Hasil penelitian ini menunjukkan bahwa kombinasi antara data mining dan teknik statistik klasik masih menjadi pendekatan yang efektif dalam analisis saham.

Metode lain yang menarik adalah perbandingan antara regresi linier dan pola historis dalam prediksi harga saham. Riyandi et al. (2023) melakukan studi komparatif antara kedua metode ini dan menemukan bahwa regresi linier memberikan hasil yang lebih akurat dibandingkan metode berbasis pola historis. Hal ini menunjukkan bahwa analisis berbasis data mining yang mengandalkan metode statistik canggih memiliki keunggulan dalam memahami pergerakan pasar secara lebih sistematis.

Penerapan data mining dalam analisis saham telah memberikan manfaat yang signifikan bagi investor dan analis keuangan. Dengan memanfaatkan teknik seperti regresi linier, LSTM, clustering, dan multiple regression, investor dapat meningkatkan akurasi prediksi dan mengurangi ketidakpastian dalam pengambilan keputusan investasi. Ke depan, integrasi antara data mining dan kecerdasan buatan diperkirakan akan semakin memperkaya strategi investasi di pasar modal.

2.3. K-Means Clustering

2.3.1. Konsep dan Prinsip K-Means Clustering

K-Means Clustering merupakan salah satu algoritma unsupervised learning yang paling banyak digunakan dalam data mining untuk mengelompokkan objek berdasarkan karakteristik yang serupa (Abrar, et. al., 2021). Algoritma ini pertama kali diperkenalkan oleh MacQueen pada tahun 1967 dan terus berkembang hingga

saat ini sebagai metode yang efisien untuk menangani berbagai jenis dataset, termasuk dalam analisis keuangan dan pasar saham. Prinsip utama dari K-Means adalah mempartisi sekumpulan data menjadi K klaster yang ditentukan sebelumnya, di mana setiap titik data akan dimasukkan ke dalam klaster dengan jarak terdekat ke pusat klaster atau centroid. Proses ini dilakukan secara iteratif hingga tidak ada perubahan signifikan dalam posisi centroid, yang menandakan bahwa klaster telah stabil.

Menurut penelitian yang dilakukan oleh Pratama et al. (2024), K-Means bekerja dengan cara menghitung jarak Euclidean antara titik data dan centroid klaster yang ada. Jarak Euclidean ini menentukan seberapa dekat suatu titik data dengan klaster tertentu, sehingga proses pengelompokan menjadi lebih optimal. Studi ini juga menyoroti bahwa salah satu tantangan dalam penerapan K-Means adalah sensitivitas terhadap pemilihan centroid awal, yang dapat berdampak pada hasil klasterisasi yang kurang optimal. Oleh karena itu, metode seperti K-Means++ sering digunakan untuk memilih centroid awal secara lebih strategis guna menghindari hasil yang bias.

Penelitian lainnya oleh Sukanto et al. (2023) membahas bagaimana K-Means sering digunakan dalam dunia keuangan untuk mengelompokkan saham berdasarkan return, volatilitas, dan faktor fundamental lainnya. Dalam konteks ini, algoritma K-Means sangat berguna dalam mengidentifikasi pola tersembunyi dalam data yang besar dan kompleks. Metode ini memungkinkan pengelompokan saham ke dalam beberapa kategori, seperti saham dengan volatilitas tinggi, saham stabil dengan return moderat, dan saham defensif yang memiliki risiko rendah.

Temuan ini menunjukkan bahwa K-Means dapat digunakan sebagai alat bantu bagi investor dalam memilih saham yang sesuai dengan strategi investasi mereka.

Selain itu, studi yang dilakukan oleh Silitonga et al. (2025) menunjukkan bahwa K-Means juga banyak digunakan dalam analisis spasial dan pengelompokan infrastruktur. Dalam penelitian mereka mengenai penyebaran Base Transceiver Station (BTS), mereka menemukan bahwa K-Means dapat digunakan untuk mengelompokkan wilayah berdasarkan kepadatan BTS. Meskipun penelitian ini tidak berfokus pada saham, hasilnya menunjukkan bahwa algoritma ini memiliki aplikasi luas dalam berbagai bidang, termasuk optimasi alokasi sumber daya berdasarkan pola data yang ditemukan.

Lebih lanjut, Sari et al. (2023) menjelaskan bahwa K-Means memiliki keunggulan dalam hal komputasi yang cepat dan skalabilitas yang tinggi, tetapi juga memiliki beberapa kelemahan, terutama dalam hal penentuan jumlah kluster yang optimal. Jika jumlah kluster tidak ditentukan dengan benar, maka hasil klusterisasi bisa menjadi kurang representatif terhadap struktur data yang sebenarnya. Untuk mengatasi masalah ini, metode seperti Elbow Method dan Silhouette Score sering digunakan untuk menentukan jumlah kluster yang ideal sebelum proses klusterisasi dilakukan.

Dari berbagai teori yang telah dikemukakan, dapat disimpulkan bahwa K-Means Clustering adalah metode yang efektif dalam mengelompokkan data ke dalam beberapa kategori berdasarkan kemiripan karakteristiknya. Algoritma ini banyak digunakan dalam berbagai bidang, termasuk analisis keuangan, pengelompokan infrastruktur, serta analisis spasial. Namun, keberhasilannya sangat bergantung pada pemilihan jumlah kluster yang tepat dan strategi inisialisasi

centroid yang baik. Oleh karena itu, meskipun K-Means adalah metode yang relatif sederhana, penerapannya memerlukan pendekatan yang hati-hati agar hasil klusterisasi dapat mencerminkan pola data yang sebenarnya.

2.3.2. Langkah-Langkah Algoritma K-Means

K-Means Clustering bekerja dalam beberapa langkah utama (Silitonga et al., 2025):

1. Menentukan jumlah klaster (K) yang optimal, Jumlah klaster dapat ditentukan menggunakan metode seperti Elbow Method atau Silhouette Score. Dalam konteks pasar saham, penelitian Pratama et al. (2024) menemukan bahwa nilai K=3 atau K=4 sering memberikan hasil optimal dalam mengelompokkan saham berdasarkan return dan volatilitasnya.
2. Menginisialisasi centroid secara acak, Centroid awal dipilih secara acak dari dataset saham yang dianalisis.
3. Menghitung jarak setiap saham ke setiap centroid, Jarak Euclidean adalah metrik yang paling umum digunakan untuk mengukur kedekatan antara saham dengan centroid suatu klaster:

$$d(x_i, c_j) = \sqrt{\sum_{k=1}^n (x_{ik} - c_{jk})^2} \quad (2.1)$$

di mana x_i adalah titik data ke-i dan c_j adalah centroid ke-j.

4. Mengelompokkan setiap saham ke centroid terdekat, Saham yang memiliki karakteristik serupa akan dikelompokkan ke dalam klaster yang sama.
5. Menghitung ulang centroid dari setiap klaster, Centroid diperbarui dengan mengambil rata-rata posisi titik-titik saham dalam klaster.

6. Mengulangi proses sampai klaster stabil, Iterasi berlanjut hingga tidak ada perubahan signifikan dalam posisi centroid, menandakan bahwa klaster telah mencapai konvergensi (Sari et al., 2023)

2.3.3. Penerapan K-Means Clustering dalam Analisis Saham

K-Means Clustering telah banyak digunakan dalam analisis pasar saham untuk membantu mengelompokkan saham berdasarkan pola yang muncul dari data historis. Metode ini memungkinkan investor dan analis untuk memahami hubungan antara saham yang memiliki karakteristik serupa serta mengidentifikasi pola yang sulit ditemukan dengan pendekatan tradisional. Menurut penelitian yang dilakukan oleh Pratama et al. (2024), K-Means dapat mengelompokkan saham berdasarkan volatilitas, return, dan rasio fundamentalnya, yang kemudian digunakan sebagai dasar dalam strategi investasi. Saham yang memiliki volatilitas tinggi sering kali membentuk klaster tersendiri, berbeda dengan saham yang lebih stabil, sehingga memudahkan investor dalam memahami potensi risiko dan peluang di pasar.

Selain itu, Sukamto et al. (2023) menunjukkan bahwa penerapan K-Means dalam analisis saham juga membantu dalam proses diversifikasi portofolio. Dengan mengelompokkan saham berdasarkan karakteristik yang serupa, investor dapat memilih saham dari klaster yang berbeda untuk mengurangi risiko sistematis. Studi ini juga menemukan bahwa saham dalam satu klaster sering kali memiliki tren harga yang serupa, sehingga jika suatu klaster menunjukkan tren positif, maka saham-saham dalam klaster tersebut berpotensi untuk mengalami kenaikan nilai. Hal ini dapat dimanfaatkan oleh investor untuk membuat keputusan yang lebih berbasis data dalam pemilihan saham untuk jangka panjang maupun jangka pendek.

Lebih lanjut, penelitian oleh Silitonga et al. (2025) menunjukkan bahwa K-Means juga digunakan dalam analisis data skala besar, termasuk dalam pengelompokan saham di indeks utama seperti LQ45. Dengan teknik clustering ini, saham dapat dikelompokkan berdasarkan likuiditas dan performa keuangan, sehingga membantu investor dalam menyusun strategi investasi yang lebih optimal. Penerapan metode ini semakin berkembang dengan adanya integrasi antara data mining dan machine learning, yang memungkinkan analisis dilakukan dalam skala besar dengan efisiensi yang lebih tinggi.

2.3.4. K-Means dan Model Markowitz dalam Optimasi Portofolio

K-Means Clustering telah banyak digunakan dalam analisis pasar saham untuk membantu mengelompokkan saham berdasarkan pola yang muncul dari data historis. Metode ini memungkinkan investor dan analis untuk memahami hubungan antara saham yang memiliki karakteristik serupa serta mengidentifikasi pola yang sulit ditemukan dengan pendekatan tradisional. Menurut penelitian yang dilakukan oleh Pratama et al. (2024), K-Means dapat mengelompokkan saham berdasarkan volatilitas, return, dan rasio fundamentalnya, yang kemudian digunakan sebagai dasar dalam strategi investasi. Saham yang memiliki volatilitas tinggi sering kali membentuk klaster tersendiri, berbeda dengan saham yang lebih stabil, sehingga memudahkan investor dalam memahami potensi risiko dan peluang di pasar.

Setelah saham dikelompokkan menggunakan K-Means Clustering, langkah berikutnya dalam optimasi portofolio adalah menerapkan Model Markowitz untuk menentukan bobot optimal dari setiap saham dalam portofolio. K-Means digunakan sebagai tahap awal dalam proses pemilihan saham dengan cara mengelompokkan saham berdasarkan return dan risiko historisnya. Dari setiap klaster yang terbentuk,

investor dapat memilih saham terbaik berdasarkan kriteria tertentu, seperti sharpe ratio atau expected return tertinggi. Pendekatan ini memungkinkan investor untuk memilih saham yang memiliki karakteristik serupa dalam satu klaster dan kemudian melakukan diversifikasi lebih lanjut dengan menggabungkan saham dari klaster yang berbeda.

Setelah tahap pemilihan saham selesai, Model Markowitz kemudian diterapkan untuk menghitung bobot optimal dari masing-masing saham dalam portofolio guna meminimalkan risiko dan memaksimalkan return. Model ini bekerja berdasarkan prinsip Mean-Variance Optimization (MVO), di mana portofolio yang terbentuk bertujuan untuk mencapai keseimbangan terbaik antara risiko dan keuntungan. Menurut penelitian yang dilakukan oleh Pratama et al. (2024), kombinasi K-Means Clustering dengan Mean-Variance Optimization dalam Model Markowitz menghasilkan portofolio yang lebih stabil dibandingkan dengan portofolio yang dipilih secara acak. Studi ini menunjukkan bahwa dengan menerapkan teknik clustering terlebih dahulu, investor dapat mengelompokkan saham dengan karakteristik risiko dan return yang serupa sebelum menentukan proporsi investasi optimal menggunakan pendekatan Markowitz.

Selain itu, penelitian oleh Sukanto et al. (2023) juga menunjukkan bahwa integrasi antara K-Means dan Model Markowitz dapat membantu investor dalam mengelola risiko lebih efektif. Dengan mengelompokkan saham berdasarkan faktor fundamental dan teknikal sebelum optimasi, proses pemilihan portofolio menjadi lebih terstruktur dan berdasarkan pola yang lebih jelas. Studi ini menegaskan bahwa strategi ini sangat berguna dalam membentuk portofolio dengan risiko terkendali

serta potensi return yang lebih baik dibandingkan strategi pemilihan saham secara konvensional.

Dengan demikian, pendekatan yang mengombinasikan K-Means Clustering dan Model Markowitz memberikan dua keuntungan utama dalam optimasi portofolio. Pertama, K-Means membantu mengurangi kompleksitas dalam pemilihan saham dengan mengelompokkan aset berdasarkan karakteristik tertentu, sehingga mempermudah investor dalam memahami perilaku masing-masing kelompok saham. Kedua, Model Markowitz memungkinkan investor untuk mendistribusikan bobot investasi secara optimal, sehingga menghasilkan portofolio dengan risiko minimal untuk tingkat return tertentu. Dengan berkembangnya teknik data mining dalam analisis keuangan, metode ini semakin menjadi pilihan utama bagi investor yang ingin mengelola portofolio mereka secara lebih ilmiah dan berbasis data.

2.4. Model Markowitz

2.4.1. Konsep Dasar Teori Portofolio Markowitz

Teori portofolio modern pertama kali diperkenalkan oleh Harry Markowitz pada tahun 1952, yang menjadi dasar bagi strategi optimasi portofolio dalam dunia keuangan. Model ini dikenal sebagai Mean-Variance Optimization (MVO), di mana investor diharapkan memilih portofolio yang memberikan return maksimum untuk tingkat risiko tertentu atau risiko minimum untuk tingkat return tertentu (Markowitz, 1952). Dalam konteks ini, risiko didefinisikan sebagai varians atau standar deviasi return portofolio, sementara return portofolio dihitung sebagai rata-rata tertimbang dari return individu tiap aset dalam portofolio.

Menurut penelitian yang dilakukan oleh Azizah et al. (2023), Model Markowitz membantu investor dalam menyeimbangkan return dan risiko dengan cara menentukan bobot optimal dari setiap aset dalam portofolio. Dengan menggunakan pendekatan ini, investor dapat membuat keputusan yang lebih berbasis data dan menghindari alokasi aset yang terlalu terkonsentrasi pada saham dengan volatilitas tinggi. Dalam studi mereka, Model Markowitz digunakan untuk mengoptimalkan portofolio saham LQ45 dengan mempertimbangkan prediksi return menggunakan metode statistik lanjutan. Hasil penelitian menunjukkan bahwa portofolio yang dioptimalkan dengan pendekatan ini memiliki risiko lebih rendah dibandingkan dengan portofolio tanpa optimasi.

Lebih lanjut, penelitian oleh Pratama et al. (2024) menunjukkan bahwa Model Markowitz dapat memberikan efisiensi portofolio yang lebih baik ketika dikombinasikan dengan metode data mining seperti K-Means Clustering. Dalam studi mereka, K-Means digunakan sebagai langkah awal untuk mengelompokkan saham berdasarkan pola return dan volatilitasnya sebelum dilakukan optimasi dengan Model Markowitz. Hasil penelitian menunjukkan bahwa strategi ini memungkinkan investor untuk memilih saham yang lebih sesuai dengan profil risiko yang diinginkan, sehingga portofolio yang terbentuk lebih terdiversifikasi dan stabil.

Selain itu, penelitian oleh Sukanto et al. (2023) menemukan bahwa pendekatan K-Means Clustering dapat meningkatkan efektivitas Model Markowitz dalam memilih saham yang memiliki karakteristik risiko dan return yang lebih seragam. Dengan mengelompokkan saham sebelum dilakukan optimasi, investor dapat menghindari pemilihan saham yang terlalu terpusat pada sektor tertentu,

sehingga membantu dalam diversifikasi risiko. Hal ini sejalan dengan prinsip dasar Model Markowitz, yaitu mengurangi risiko portofolio melalui diversifikasi aset yang tidak berkorelasi tinggi satu sama lain.

Berdasarkan berbagai penelitian yang telah dikaji, dapat disimpulkan bahwa Model Markowitz merupakan salah satu pendekatan paling efektif dalam optimasi portofolio saham, terutama ketika dikombinasikan dengan metode K-Means Clustering. Dengan memanfaatkan teknik clustering sebelum optimasi portofolio dilakukan, investor dapat mengidentifikasi saham yang memiliki karakteristik return dan risiko yang serupa, sehingga proses alokasi aset menjadi lebih optimal. Oleh karena itu, pendekatan ini semakin banyak digunakan dalam dunia investasi modern untuk membangun portofolio yang lebih terdiversifikasi dan efisien dalam pengelolaan risiko.

2.4.2. Expected Return, Variance, dan Covariance dalam Portofolio

Dalam teori portofolio Markowitz, konsep expected return, variance, dan covariance merupakan elemen utama dalam pengukuran performa dan risiko portofolio. Expected return digunakan untuk mengestimasi keuntungan yang diharapkan dari suatu portofolio, sementara variance dan covariance digunakan untuk mengukur risiko serta hubungan antar aset dalam portofolio (Markowitz, 1952).

1. Expected Return dalam Portofolio

Expected return atau return yang diharapkan adalah rata-rata tertimbang dari return individual setiap aset dalam portofolio. Perhitungan expected return portofolio ($E(R_p)$) diberikan oleh rumus berikut:

$$E(R_p) = \sum_{i=1}^n w_i E(R_i) \quad (2.2)$$

di mana:

$(E(R_p))$ = expected return portofolio

w_i = proporsi bobot investasi pada aset ke-i

$(E(R_i))$ = expected return dari aset ke-i

n = jumlah aset dalam portofolio

Menurut penelitian yang dilakukan oleh Azizah et al. (2023), expected return menjadi indikator utama dalam optimasi portofolio karena nilai ini mencerminkan tingkat keuntungan rata-rata yang dapat diperoleh investor berdasarkan komposisi aset dalam portofolio. Dengan memilih aset yang memiliki return yang lebih tinggi dalam portofolio, investor dapat meningkatkan potensi keuntungan mereka tanpa harus meningkatkan risiko secara proporsional.

2. Variance dalam Portofolio

Variance (σ^2) mengukur tingkat dispersi return aset dari nilai rata-ratanya, yang digunakan untuk menilai risiko yang melekat pada suatu investasi. Variance portofolio tidak hanya bergantung pada variance individual aset tetapi juga pada hubungan antar aset dalam portofolio (Markowitz, 1952). Variance dari suatu portofolio dihitung dengan rumus berikut:

$$\sigma_p^2 = \sum_{i=1}^n w_i^2 \sigma_i^2 + \sum_{i=1}^n \sum_{j=1}^n w_i w_j \sigma_{ij} \quad (2.3)$$

di mana:

σ_p^2 = variance portofolio

W_i dan W_j = bobot aset ke-i dan ke-j

σ_i^2 = variance dari aset ke-i

σ_{ij}^2 = covariance antara aset ke-i dan ke-j

Penelitian oleh Pratama et al. (2024) menunjukkan bahwa variance sangat berpengaruh dalam pemilihan portofolio optimal. Dengan memahami variance dari setiap aset, investor dapat mengelola risiko secara lebih efektif dan menyeimbangkan proporsi aset yang memiliki tingkat volatilitas tinggi dengan aset yang lebih stabil.

3. Covariance dan Korelasi Antar Aset dalam Portofolio

Covariance σ_{ij} mengukur hubungan antara return dua aset dalam portofolio. Jika covariance bernilai positif, berarti kedua aset bergerak searah, sedangkan covariance negatif menunjukkan bahwa kedua aset bergerak berlawanan arah. Formula untuk covariance adalah:

$$\sigma_{ij} = E[(R_i - E(R_i))(R_j - E(R_j))] \quad (2.4)$$

Untuk memahami hubungan antar aset lebih lanjut, covariance sering dikonversi menjadi koefisien korelasi (ρ_{ij}), yang dihitung dengan:

$$\rho_{ij} = \frac{\sigma_{ij}}{\sigma_i \sigma_j} \quad (2.5)$$

Menurut Sukanto et al. (2023), korelasi antar aset memainkan peran penting dalam diversifikasi risiko dalam portofolio. Jika investor memilih aset yang memiliki korelasi rendah atau negatif, risiko portofolio dapat dikurangi secara signifikan tanpa harus mengorbankan return yang diharapkan. Studi ini menekankan bahwa dalam praktiknya, investor sebaiknya memilih saham dengan tingkat korelasi yang rendah untuk memaksimalkan manfaat diversifikasi.

2.4.3. Efisien Frontier dan Optimalisasi Risiko-Return

a. Efisien Frontier dalam Model Markowitz

Dalam teori portofolio yang diperkenalkan oleh Harry Markowitz (1952), efisien frontier merupakan kumpulan portofolio optimal yang menawarkan return

maksimum untuk tingkat risiko tertentu atau risiko minimum untuk tingkat return tertentu. Portofolio yang berada di bawah efisien frontier dianggap tidak efisien karena ada portofolio lain dengan return lebih tinggi pada tingkat risiko yang sama atau dengan risiko lebih rendah untuk return yang sama.

Secara matematis, efisien frontier dibentuk dengan memplot kombinasi berbagai portofolio pada grafik dengan sumbu horizontal mewakili risiko (standar deviasi) dan sumbu vertikal mewakili return yang diharapkan. Kurva yang terbentuk menunjukkan trade-off antara risiko dan return, di mana setiap titik pada kurva tersebut mewakili portofolio yang efisien.

b. Optimalisasi Risiko-Return

Optimalisasi risiko-return bertujuan untuk menemukan kombinasi aset yang memberikan return yang diharapkan tertinggi dengan tingkat risiko yang dapat diterima atau, sebaliknya, meminimalkan risiko untuk tingkat return tertentu. Proses ini melibatkan penentuan bobot masing-masing aset dalam portofolio sehingga portofolio tersebut berada pada efisien frontier.

Penelitian oleh Pratama et al. (2024) menunjukkan bahwa penggunaan K-Means Clustering dalam tahap awal pemilihan saham dapat meningkatkan efektivitas optimalisasi portofolio. Dengan mengelompokkan saham berdasarkan karakteristik return dan risiko historisnya, investor dapat memilih perwakilan saham dari setiap kluster untuk membentuk portofolio yang lebih terdiversifikasi. Setelah itu, Model Markowitz diterapkan untuk menentukan bobot optimal dari setiap saham dalam portofolio, sehingga mencapai keseimbangan risiko-return yang optimal.

Selain itu, penelitian oleh Gubu et al. (2021) menggunakan metode K-Medoids Clustering dengan ukuran jarak Dynamic Time Warping (DTW) untuk mengelompokkan saham berdasarkan pola pergerakan harganya. Saham yang terpilih dari setiap klaster kemudian digunakan untuk membentuk portofolio optimal menggunakan model portofolio Mean-Variance klasik. Hasil penelitian menunjukkan bahwa pendekatan ini dapat meningkatkan kinerja portofolio dengan Sharpe ratio yang lebih tinggi dibandingkan dengan metode tradisional.

Dengan demikian, integrasi teknik clustering seperti K-Means atau K-Medoids dalam proses pembentukan portofolio dapat membantu investor dalam mengidentifikasi kelompok saham dengan karakteristik serupa, yang selanjutnya mempermudah proses optimalisasi risiko-return menggunakan Model Markowitz. Pendekatan ini tidak hanya meningkatkan efisiensi dalam pemilihan saham tetapi juga berkontribusi pada pembentukan portofolio yang lebih stabil dan sesuai dengan preferensi risiko investor.

2.4.4. Integrasi dengan K-Means Clustering dalam Pemilihan Saham

Dalam teori portofolio yang diperkenalkan oleh Harry Markowitz (1952), efisien frontier merupakan kumpulan portofolio optimal yang menawarkan return maksimum untuk tingkat risiko tertentu atau risiko minimum untuk tingkat return tertentu. Portofolio yang berada di bawah efisien frontier dianggap tidak efisien karena ada portofolio lain dengan return lebih tinggi pada tingkat risiko yang sama atau dengan risiko lebih rendah untuk return yang sama.

Dalam konteks optimasi portofolio saham, integrasi antara K-Means Clustering dengan Model Markowitz telah menjadi pendekatan yang semakin banyak digunakan untuk meningkatkan efisiensi diversifikasi dan manajemen

risiko. Metode ini menggabungkan analisis klasterisasi saham berdasarkan karakteristik tertentu sebelum dilakukan optimasi bobot aset dalam portofolio menggunakan pendekatan Mean-Variance Optimization (MVO) dari Model Markowitz.

Menurut penelitian yang dilakukan oleh Abrar (2021), penggunaan K-Means Clustering dalam pemilihan saham memungkinkan investor untuk mengelompokkan saham berdasarkan risiko dan return historisnya, sehingga diversifikasi dapat dilakukan dengan lebih optimal. Dalam studi tersebut, saham-saham syariah yang termasuk dalam Indeks Saham Syariah Indonesia (ISSI) dikelompokkan menjadi enam klaster berdasarkan return dan volatilitasnya selama pandemi COVID-19. Hasil klasterisasi menunjukkan bahwa saham yang berasal dari klaster berbeda memiliki karakteristik risiko yang lebih bervariasi, yang kemudian digunakan sebagai dasar dalam pembentukan portofolio yang lebih terdiversifikasi (Abrar, 2021).

Lebih lanjut, penelitian oleh Pratama et al. (2024) juga menemukan bahwa pemanfaatan K-Means sebelum optimasi portofolio dengan Model Markowitz dapat meningkatkan stabilitas investasi. Dengan mengelompokkan saham berdasarkan karakteristik fundamental dan teknikalnya, investor dapat memilih perwakilan saham dari setiap klaster untuk mengurangi korelasi antar aset dalam portofolio, sehingga risiko dapat ditekan tanpa mengorbankan potensi return.

Studi oleh Sukanto et al. (2023) menambahkan bahwa pendekatan ini tidak hanya efektif dalam pasar konvensional, tetapi juga dalam pasar saham berbasis syariah, di mana pemilihan saham harus memenuhi prinsip-prinsip investasi syariah. Dengan menggunakan K-Means untuk mengelompokkan saham

berdasarkan indikator return, volatilitas, dan Sharpe ratio, hasil penelitian menunjukkan bahwa portofolio yang dibentuk memiliki return lebih tinggi dibandingkan dengan indeks pasar syariah secara keseluruhan.

Dalam penerapannya, integrasi antara K-Means Clustering dan Model Markowitz umumnya dilakukan dengan beberapa langkah sebagai berikut:

1. Pengelompokan saham berdasarkan return dan risiko historis menggunakan K-Means, Saham dengan karakteristik yang serupa dikelompokkan dalam satu klaster. Proses ini membantu menghindari pemilihan saham yang memiliki pola pergerakan harga yang terlalu mirip, sehingga meningkatkan efektivitas diversifikasi.
2. Pemilihan saham terbaik dari setiap klaster berdasarkan indikator kinerja. Setelah klasterisasi, investor memilih saham terbaik dari setiap klaster berdasarkan indikator seperti Sharpe Ratio atau Expected Return.

Pendekatan ini memastikan bahwa hanya saham dengan potensi return tertinggi dan risiko terkendali yang masuk dalam portofolio.

3. Optimasi bobot saham dalam portofolio menggunakan Model Markowitz, Setelah pemilihan saham dilakukan, langkah selanjutnya adalah menghitung kombinasi bobot optimal dari setiap saham dalam portofolio. Model Markowitz digunakan untuk meminimalkan risiko sambil memaksimalkan return, berdasarkan matriks varians-kovarians antara aset-aset yang terpilih.

Menurut penelitian oleh Tohendry & Jollyta (2023), pendekatan ini terbukti lebih unggul dibandingkan dengan metode pemilihan saham secara acak atau

berbasis sektor saja. Mereka menemukan bahwa portofolio yang dibangun dengan teknik ini memiliki risk-adjusted return yang lebih tinggi, yang berarti tingkat keuntungan yang diperoleh lebih seimbang dengan risiko yang diambil.

Integrasi K-Means Clustering dalam pemilihan saham tidak hanya meningkatkan kualitas portofolio dari segi return dan risiko, tetapi juga membantu investor dalam membuat keputusan yang lebih berbasis data. Dengan memanfaatkan teknik ini, investor dapat menghindari overexposure pada saham dengan korelasi tinggi, sehingga dampak volatilitas pasar terhadap portofolio dapat diminimalkan.

2.5. Python dalam Analisis Data Keuangan

Python telah menjadi salah satu bahasa pemrograman utama dalam analisis data keuangan karena fleksibilitasnya serta dukungan pustaka yang luas untuk pemrosesan data, analisis statistik, dan implementasi machine learning (McKinney, 2017). Dalam beberapa tahun terakhir, penerapan Python dalam analisis data keuangan semakin meningkat, terutama dalam pengolahan data saham, optimasi portofolio, serta prediksi harga saham menggunakan machine learning (VanderPlas, 2016).

Beberapa pustaka utama yang sering digunakan dalam analisis data keuangan dengan Python meliputi Pandas, NumPy, Matplotlib, dan Scikit-Learn. Pustaka-pustaka ini memungkinkan investor dan analis keuangan untuk mengakses, membersihkan, dan memvisualisasikan data keuangan dengan lebih efisien, sehingga pengambilan keputusan dapat lebih berbasis data.

2.5.1. Pustaka Python yang Digunakan

Dalam penelitian ini, beberapa pustaka Python yang digunakan antara lain:

a. Pandas

Pandas adalah pustaka Python yang dirancang khusus untuk manipulasi dan analisis data berbasis tabel seperti spreadsheet atau database relasional (McKinney, 2017). Dalam konteks keuangan, Pandas sering digunakan untuk memproses data historis harga saham, laporan keuangan, dan data time-series lainnya. Struktur data utama dalam Pandas adalah DataFrame, yang memungkinkan manipulasi data dalam bentuk tabel dua dimensi dengan dukungan fitur seperti penggabungan data, filtering, serta analisis statistik dasar.

b. NumPy

NumPy adalah pustaka yang berfungsi untuk melakukan komputasi numerik dan operasi array multidimensi secara efisien (Harris et al., 2020). Pustaka ini digunakan dalam analisis keuangan untuk berbagai keperluan, seperti perhitungan statistik return saham, volatilitas, serta korelasi antar aset dalam portofolio. NumPy juga menjadi dasar bagi banyak pustaka lain, termasuk Pandas dan Scikit-Learn, dalam menangani perhitungan berbasis array yang lebih kompleks.

c. Matplotlib

Matplotlib adalah pustaka visualisasi data yang digunakan untuk membuat grafik dan diagram dalam analisis data keuangan. Dengan Matplotlib, analisis keuangan dapat memvisualisasikan pergerakan harga saham, distribusi return, serta plotting efisien frontier dalam optimasi portofolio menggunakan Model Markowitz.

Penggunaan grafik dalam analisis data keuangan sangat penting untuk memahami pola tren serta hubungan antar variabel dalam dataset.

d. Scikit-Learn

Scikit-Learn adalah pustaka machine learning yang banyak digunakan dalam data mining dan prediksi harga saham. Pustaka ini mencakup berbagai algoritma pembelajaran mesin, termasuk K-Means Clustering untuk pengelompokan saham dan regresi untuk memprediksi return saham berdasarkan faktor historis. Penggunaan Scikit-Learn dalam kombinasi dengan K-Means dan Model Markowitz dapat meningkatkan akurasi optimasi portofolio serta mengurangi risiko investasi.

2.5.2. Implementasi Data Mining dengan Python

Python digunakan untuk mengotomatisasi proses pengolahan data saham, clustering saham berdasarkan K-Means, dan optimasi portofolio dengan Model Markowitz. Dengan menggunakan kombinasi pustaka yang disebutkan di atas, analisis dapat dilakukan dengan efisiensi dan akurasi tinggi.

Penerapan data mining dalam analisis data keuangan menggunakan Python mencakup beberapa langkah utama yang saling berhubungan (McKinney, 2017):

a. Pengumpulan Data

Data keuangan dapat diperoleh dari berbagai sumber seperti Yahoo Finance, Bloomberg, atau database keuangan lainnya dalam bentuk CSV atau melalui API. Pandas digunakan untuk membaca dan mengelola data yang diperoleh.

b. Preprocessing Data

Melakukan pembersihan data, menangani nilai yang hilang (missing values), dan menormalkan data agar siap digunakan dalam model machine learning

(Harris et al., 2020). NumPy dan Pandas membantu dalam transformasi data sebelum analisis lebih lanjut dilakukan.

c. Eksplorasi dan Visualisasi Data

Matplotlib digunakan untuk memvisualisasikan tren harga saham, volatilitas, dan pergerakan portofolio. Teknik plotting dapat digunakan untuk memahami korelasi antar saham serta efisien frontier dalam Model Markowitz.

d. Pemodelan dan Analisis Data

Scikit-Learn digunakan untuk menerapkan algoritma machine learning seperti K-Means untuk clustering saham dan regresi linier untuk memprediksi harga saham berdasarkan faktor fundamental. Model prediktif dapat dikembangkan untuk mengidentifikasi saham yang memiliki potensi return tinggi.

e. Evaluasi dan Optimasi Portofolio

Setelah model diterapkan, hasilnya dievaluasi dengan metrik seperti Mean Absolute Error (MAE) atau R-Squared untuk regresi, serta Silhouette Score untuk clustering. Jika digunakan dalam konteks optimasi portofolio, Model Markowitz dapat diterapkan untuk menghitung bobot optimal dari saham yang telah dipilih menggunakan K-Means Clustering.

2.6. Penelitian Terdahulu

Berikut adalah beberapa penelitian yang relevan dengan penelitian ini, yang mengkaji penggunaan K-Means Clustering dan Model Markowitz dalam optimasi portofolio saham.

Tabel 2.1. Penelitian Terdahulu

No	Penulis (Tahun)	Judul Penelitian	Hasil Penelitian
1	Ali Sulthan Abrar (2024)	Diversifikasi Portofolio Saham Syariah Indonesia Menggunakan Algoritma K-Means Clustering: Studi Kasus Pandemi Covid-19	Penggunaan metode klasterisasi dinilai efektif dalam mengidentifikasi sifat-sifat pergerakan harga. Dengan menggunakan metode klasterisasi, ditemukan pada pasar modal syariah terdapat 6 klaster yang masing-masing memiliki pergerakan harga yang berbeda. Investor dapat menggunakan hasil klasterisasi untuk membuat keputusan investasi yang lebih baik seperti memilih klaster yang sifat pergerakan harganya sesuai dengan target return dan profil resiko investor.
2	Anggi Srimurdianti Sukamto, Wawan Setiawan, Enda Esyudha Pratama (2023)	Data Mining untuk Pengelompokan Saham pada Sektor Energi dengan Metode K-Means	Penelitian ini menggunakan K-Means Clustering untuk mengelompokkan saham di sektor energi berdasarkan karakteristik fundamental. Hasil clustering menunjukkan bahwa algoritma ini dapat membantu investor dalam memilih saham yang sesuai dengan preferensi investasi
3	Derrick Tohendry, Deny Jollyta (2023)	Penerapan Algoritma K-Means Clustering Untuk Pengelompokan Saham Berdasarkan Price Earning Ratio Dan Price To Book Value	Penelitian ini menerapkan K-Means Clustering untuk mengelompokkan saham berdasarkan rasio keuangan PER dan PBV, dengan hasil pengelompokan yang dapat membantu investor dalam seleksi saham berdasarkan valuasi
4	Yogi Pratama, Evy Sulistianingsih, Naomi Nessyana Debataraaja, Nurfitri Imro'ah (2024)	K-Means Clustering dan Mean Variance Efficient Portfolio dalam Portofolio Saham	Penelitian ini menggunakan K-Means Clustering dalam memilih 11 saham dari LQ45 dan MVEP untuk diversifikasi. Hasilnya menunjukkan bahwa clustering dapat meningkatkan efektivitas pemilihan saham dan MVEP mampu melakukan diversifikasi portofolio
5	Sunariyah (2022)	Pengaruh Model Mean-Variance dalam Optimasi Portofolio Saham	Studi ini menyoroti kelemahan Model Markowitz, khususnya asumsi tentang korelasi aset yang tetap. Hasil penelitian merekomendasikan penggunaan teknik tambahan seperti clustering atau machine learning untuk mengatasi keterbatasan model ini

Berdasarkan kajian penelitian terdahulu, berbagai studi telah menunjukkan efektivitas metode K-Means Clustering dan Model Markowitz dalam optimasi portofolio saham. Penelitian oleh Sukamto et. al., (2023) menunjukkan bahwa K-

Means Clustering efektif dalam mengelompokkan saham di sektor energi berdasarkan karakteristik fundamental. Hasilnya menunjukkan bahwa pengelompokan ini dapat membantu investor dalam memilih saham yang sesuai dengan preferensi investasi mereka. Temuan serupa juga dikonfirmasi oleh Tohendry dan Jollyta (2023) yang menerapkan K-Means Clustering berdasarkan rasio keuangan seperti Price-to-Earnings Ratio (PER) dan Price-to-Book Value (PBV). Studi ini menyimpulkan bahwa pendekatan clustering dapat membantu dalam seleksi saham berbasis valuasi fundamental, memberikan alternatif yang lebih sistematis dibandingkan pemilihan saham secara subjektif.

Selain itu, penelitian yang dilakukan oleh Tohendry (2023) menyoroti pentingnya K-Means Clustering dalam pemilihan saham di Bursa Efek Indonesia (IDX). Penelitian ini menunjukkan bahwa penggunaan teknik clustering mampu meningkatkan efektivitas pemilihan saham dibandingkan dengan metode acak, terutama dalam menciptakan portofolio yang lebih terdiversifikasi. Sementara itu, penelitian oleh Abrar menunjukkan bahwa metode klasterisasi dapat mengidentifikasi sifat pergerakan harga saham syariah selama pandemi Covid-19. Hasil penelitian ini menemukan bahwa terdapat enam klaster yang memiliki pergerakan harga yang berbeda-beda, sehingga dapat membantu investor dalam membuat keputusan investasi yang lebih baik sesuai dengan target return dan profil risiko mereka.

Di sisi lain, penelitian oleh Sunariyah (2022) membahas kelemahan Model Markowitz dalam menangani asumsi tentang korelasi aset yang tetap. Studi ini merekomendasikan bahwa model ini dapat ditingkatkan dengan teknik tambahan, seperti clustering atau machine learning, guna mengatasi keterbatasan dalam

asumsi hubungan linear antar aset dalam portofolio. Hal ini menunjukkan bahwa meskipun Model Markowitz merupakan standar dalam optimasi portofolio, ada kebutuhan untuk mengombinasikannya dengan pendekatan modern seperti data mining untuk meningkatkan keakuratan dan efektivitasnya dalam pasar yang dinamis.

Berdasarkan penelitian-penelitian sebelumnya, dapat disimpulkan bahwa kombinasi K-Means Clustering dan Model Markowitz merupakan pendekatan yang menjanjikan dalam optimasi portofolio saham. K-Means membantu dalam seleksi saham berdasarkan kesamaan karakteristik fundamental dan teknikal, sementara Model Markowitz digunakan untuk mengoptimalkan bobot portofolio berdasarkan return dan risiko. Penelitian ini akan mengonfirmasi efektivitas pendekatan ini dalam konteks pasar saham Indonesia, khususnya pada Bursa Efek Indonesia (IDX), dengan tujuan untuk menghasilkan portofolio optimal yang lebih terdiversifikasi dan efisien.

BAB III

METODOLOGI PENELITIAN

3.1. Jenis Penelitian

Penelitian ini merupakan penelitian kuantitatif deskriptif yang bertujuan untuk menganalisis pola dan karakteristik saham menggunakan K-Means Clustering serta mengoptimalkan portofolio dengan Model Markowitz. Data yang digunakan dalam penelitian ini bersifat sekunder, yang diperoleh dari Yahoo Finance dengan rentang waktu 1 Januari 2019 – 31 Desember 2024.

Metode kuantitatif deskriptif digunakan karena penelitian ini tidak melakukan intervensi langsung terhadap pasar saham, melainkan menganalisis data yang telah ada untuk mendapatkan pola optimal dalam pemilihan saham dan komposisi portofolio. Melalui pendekatan ini, penelitian akan menghasilkan gambaran pola clustering saham serta efektivitas Model Markowitz dalam mengoptimalkan portofolio berdasarkan hasil clustering.

Penelitian ini juga bersifat data-driven, di mana teknik machine learning (K-Means Clustering) digunakan untuk mengelompokkan saham berdasarkan karakteristik tertentu, sedangkan Model Markowitz diterapkan untuk menentukan komposisi portofolio yang optimal berdasarkan risiko dan return. Dengan demikian, hasil dari penelitian ini dapat memberikan wawasan bagi investor dalam membangun strategi investasi berbasis data.

3.2. Sumber Data dan Teknik Pengumpulan Data

3.2.1. Sumber Data

Data yang digunakan dalam penelitian ini berasal dari Yahoo Finance, yang merupakan salah satu sumber data keuangan terbesar yang menyediakan informasi harga saham, volume perdagangan, rasio keuangan, dan data teknikal lainnya. Data diambil menggunakan Yahoo Finance API melalui pustaka `yfinance` di Python, yang memungkinkan akses langsung ke data historis saham yang terdaftar di Bursa Efek Indonesia (IDX).

Jenis data yang digunakan dalam penelitian ini meliputi:

1. Data Fundamental

- 1) Price-to-Earnings Ratio (P/E): Menunjukkan valuasi saham berdasarkan laba perusahaan.
- 2) Price-to-Book Value (PBV): Menilai harga saham dibandingkan dengan nilai buku perusahaan.
- 3) Return on Equity (ROE): Mengukur efisiensi penggunaan ekuitas dalam menghasilkan keuntungan.

Tabel 3.1. Data Fundamental

Kode	PE	PBV	ROE
AKRA	10.52	2.02	16.74
AMRT	27.6	5.3	19.28
ARTO	207.34	3.12	1.52
BBCA	18.86	4.04	21.71
BBNI	7.29	0.96	13.46
...	...	...	...
MEDC	4.27	12470.59	17.44
SCMA	21.19	1.69	5.76
SMGR	24.29	0.4	1.61
TBIG	32.76	4.46	12.41
UNVR	21.36	31.06	121.82

2. Data Teknikal

1) Harga saham historis: Close.

Tabel 3.2. Harga Close Saham

Date	AKRA	AMRT	ARTO	...	SMGR	TBIG	UNVR
1/1/2019	627.7768	875.6530	22.959659	...	10064.371	619.03820	7324.098
1/2/2019	620.46008	870.97045	23.458782	...	10042.490	610.44036	7485.423
1/3/2019	623.38665	866.28784	22.710098	...	10086.249	617.31860	7662.879
1/4/2019	636.55694	861.60516	22.834878	...	10064.371	617.31860	7711.275
1/7/2019	680.4573	852.2398	22.21097	...	9976.8544	617.31860	7840.336
...	...	...	...	...	...	...	...
12/20/2024	1140	2780	2240	...	3260	1900	1750
12/23/2024	1130	2940	2400	...	3350	1910	1785
12/24/2024	1120	2780	2390	...	3240	1955	1780
12/27/2024	1120	2800	2360	...	3300	2150	1840
12/30/2024	1120	2850	2430	...	3290	2100	1885

2) Return saham historis: Perubahan harga dalam periode tertentu.

3) Volatilitas saham: Mengukur seberapa besar pergerakan harga saham dalam periode tertentu.

3.2.2. Teknik Pengumpulan Data

Data akan dikumpulkan menggunakan API Yahoo Finance yang diakses melalui pustaka yfinance dalam Python. Proses pengambilan data dilakukan dengan langkah-langkah berikut:

1. Mengimpor pustaka yfinance dalam Python.
2. Menentukan daftar kode saham IDX yang akan dianalisis.
3. Mengambil data harga saham historis dengan interval waktu yang telah ditentukan.
4. Menghitung indikator teknikal dan fundamental berdasarkan data yang diperoleh.
5. Menyimpan data dalam format yang siap digunakan untuk analisis lebih lanjut.

Data yang telah dikumpulkan kemudian akan digunakan dalam K-Means Clustering untuk pengelompokan saham dan Model Markowitz untuk optimasi portofolio.

3.3. Teknik Analisis Data

Teknik analisis data dalam penelitian ini terdiri dari beberapa tahapan, yaitu preprocessing data, penerapan K-Means Clustering, optimasi portofolio dengan Model Markowitz, dan evaluasi portofolio yang terbentuk. Seluruh proses dilakukan menggunakan bahasa pemrograman Python, dengan pustaka seperti Pandas, NumPy, Scikit-Learn, dan PyPortfolioOpt.

3.3.1. Preprocessing Data

Sebelum dilakukan analisis, data yang diperoleh dari Yahoo Finance API akan melalui tahap preprocessing untuk memastikan data bersih dan siap digunakan dalam analisis. Langkah-langkah preprocessing meliputi:

1. Menghapus data yang hilang (missing values) dan mengganti dengan interpolasi atau metode lain.
2. Normalisasi data untuk memastikan semua variabel berada dalam skala yang seragam sebelum diterapkan ke dalam K-Means Clustering.
3. Menghitung return saham harian dan volatilitas berdasarkan harga saham historis.

3.3.2. Penerapan K-Means Clustering

Algoritma K-Means Clustering digunakan dalam penelitian ini untuk mengelompokkan saham berdasarkan karakteristik fundamental dan teknikalnya. Algoritma ini bekerja secara iteratif dengan memperbarui centroid hingga tidak ada

perubahan signifikan dalam pembagian kluster. Langkah-langkah utama dalam algoritma K-Means adalah sebagai berikut:

1. Menentukan Jumlah Kluster

Sebelum melakukan proses clustering, nilai K harus ditentukan terlebih dahulu. Penelitian ini menggunakan metode Elbow Method untuk menentukan nilai K yang optimal.

2. Menginisialisasi Centroid Awal

Centroid awal dipilih secara acak dari dataset saham yang dianalisis. Pemilihan centroid yang baik dapat meningkatkan stabilitas hasil clustering.

3. Menghitung Jarak Saham ke Centroid Terdekat

Setiap saham dihitung jaraknya ke setiap centroid menggunakan Jarak Euclidean, dengan rumus:

$$d(x, c) = \sqrt{\sum_{i=1}^n (x_i - c_i)^2} \quad (3.1)$$

Di mana:

$d(x, c)$ = jarak antara titik data x (saham) dan centroid c

x_i = nilai fitur ke-i dari saham

c_i = nilai fitur ke-i dari centroid kluster

n = jumlah fitur dalam dataset

Saham akan dikelompokkan ke dalam kluster dengan centroid terdekat.

4. Memperbarui Centroid Setiap Kluster

Setelah semua saham dikelompokkan, centroid baru dari masing-masing kluster dihitung dengan mengambil rata-rata seluruh titik dalam kluster:

$$c_j = \frac{1}{N} \sum_{i=1}^N x_i \quad (3.2)$$

Di mana:

c_j = centroid baru dari klaster j

N = jumlah saham dalam klaster

x_i = nilai fitur dari saham dalam klaster tersebut

5. Mengulangi Proses hingga Konvergensi

Langkah 3 dan 4 diulang hingga centroid tidak lagi berubah secara signifikan atau jumlah iterasi maksimum tercapai. Algoritma ini dapat berhenti jika:

- a. Tidak ada saham yang berpindah klaster antara iterasi sebelumnya dan saat ini.
- b. Perubahan posisi centroid sudah sangat kecil.

6. Evaluasi Hasil Clustering

Untuk mengukur seberapa baik hasil clustering, digunakan Silhouette Score, dengan rumus:

$$S(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))} \quad (3.3)$$

Di mana:

$S(i)$ = nilai silhouette untuk saham ke- i

$a(i)$ = rata-rata jarak saham ke semua saham dalam klasternya sendiri

$b(i)$ = rata-rata jarak saham ke klaster terdekat lainnya

Semakin tinggi nilai Silhouette Score (mendekati 1), semakin baik kualitas klasterisasi.

Saham yang telah dikelompokkan berdasarkan K-Means Clustering akan digunakan sebagai dasar dalam pemilihan saham untuk portofolio yang akan dioptimalkan menggunakan Model Markowitz.

3.3.3. Pemilihan Saham dari Setiap Klaster

Setelah proses pengelompokan saham dilakukan menggunakan algoritma K-Means Clustering dan diperoleh hasil klasterisasi dengan jumlah klaster optimal, langkah selanjutnya adalah melakukan seleksi terhadap saham yang akan dimasukkan ke dalam portofolio.

Pemilihan saham dilakukan berdasarkan nilai return rata-rata tahunan (mean return) tertinggi dari setiap klaster. Tujuannya adalah untuk mewakili setiap klaster dengan saham yang memiliki potensi keuntungan terbaik berdasarkan kinerja historisnya. Pemilihan ini dilakukan secara otomatis menggunakan program Python, dengan logika sebagai berikut:

1. Data saham hasil clustering disimpan dalam satu DataFrame (`clustered_df`) yang mencakup fitur teknikal dan fundamental, serta label klaster.
2. Untuk setiap klaster, dicari indeks saham dengan nilai mean return tertinggi.
3. Hasil seleksi berupa saham terpilih, masing-masing satu dari tiap klaster, yang selanjutnya digunakan sebagai input dalam proses optimasi portofolio.

Metode seleksi ini memastikan bahwa setiap klaster direpresentasikan oleh saham yang memiliki performa historis terbaik, sekaligus mempertahankan diversifikasi antar segmen pasar sesuai hasil pengelompokan sebelumnya.

3.3.4. Optimasi Portofolio dengan Model Markowitz

Setelah proses seleksi saham dari masing-masing klaster, tahap berikutnya adalah optimasi bobot investasi untuk membentuk portofolio yang efisien. Optimasi dilakukan menggunakan pendekatan Model Markowitz dengan tujuan memaksimalkan rasio Sharpe.

Model ini menggabungkan rata-rata return tahunan dan kovarian antar saham untuk menghasilkan komposisi portofolio dengan trade-off optimal antara return dan risiko.

Adapun proses optimasi meliputi:

1. Penghitungan Mean Return dan Kovarian Return

Rata-rata return tahunan dan kovarian antar return saham dihitung berdasarkan data historis harga saham periode Januari 2019 hingga Desember 2024.

2. Formulasi Fungsi Objektif

Fungsi objektif yang digunakan adalah negatif Sharpe Ratio, untuk kemudian diminimalkan:

$$\text{Sharpe Ratio} = \frac{R_p - R_f}{\sigma_p} \quad (3.4)$$

dengan:

R_p = return portofolio

R_f = risk-free rate (diasumsikan sebesar 3% per tahun)

σ_p = volatilitas portofolio

3. Constraint dan Batasan

- a. Jumlah bobot portofolio harus sama dengan 1 ($\sum w_i = 1$).
- b. Tidak diperbolehkan short-selling (bobot setiap saham ≥ 0).

4. Metode Optimasi

Optimasi dilakukan menggunakan metode Sequential Least Squares Programming (SLSQP) melalui pustaka `scipy.optimize.minimize` di Python.

3.3.5. Evaluasi Portofolio

Setelah portofolio optimal terbentuk, dilakukan proses evaluasi performa untuk menilai efektivitas strategi yang digunakan. Evaluasi dilakukan melalui:

1. Simulasi Backtest Tahun 2024

Portofolio dengan bobot optimal diuji pada data harga saham aktual selama periode 1 Januari 2024 hingga 31 Desember 2024. Perhitungan return kumulatif dilakukan dengan modal awal asumsi 100 untuk mengukur pertumbuhan nilai portofolio.

2. Penghitungan Return Total

Return total portofolio selama periode backtest dihitung menggunakan formula:

$$Total\ Return = \left(\frac{Nilai\ Akhir}{Nilai\ Awal} - 1 \right) \times 100\% \quad (3.5)$$

3. Perbandingan dengan Benchmark (Indeks LQ45)

Untuk mengukur keunggulan relatif portofolio, hasil kinerja dibandingkan dengan return indeks LQ45 pada periode yang sama. Indeks ini diunduh menggunakan simbol ^JKLQ45 dari Yahoo Finance dan dinormalisasi ke basis 100.

4. Analisis Grafik Kinerja

Grafik pertumbuhan nilai portofolio dan indeks LQ45 digunakan untuk melihat visualisasi performa relatif sepanjang tahun 2024.

Melalui evaluasi ini, dapat dianalisis apakah portofolio hasil optimasi mengungguli performa pasar atau tidak, baik dari sisi return maupun kestabilan fluktuasi nilai.

3.4. Alat dan Perangkat Lunak

Penelitian ini dilakukan dengan menggunakan pemrograman Python, yang memiliki berbagai pustaka yang mendukung analisis data, machine learning, dan optimasi portofolio saham. Seluruh proses pengolahan data, penerapan algoritma K-Means Clustering, serta optimasi portofolio dengan Model Markowitz akan dijalankan di Google Colab.

3.4.1. Google Colab sebagai Lingkungan Pengembangan

Google Colab dipilih sebagai lingkungan pengembangan karena memiliki beberapa keunggulan, yaitu:

1. Berbasis cloud, sehingga tidak memerlukan instalasi perangkat lunak tambahan.
2. Mendukung pustaka Python secara default, seperti NumPy, Pandas, Scikit-Learn, dan Matplotlib.
3. Memiliki dukungan GPU, yang dapat mempercepat proses komputasi jika diperlukan.
4. Mudah digunakan dan dapat diakses dari mana saja tanpa perlu konfigurasi khusus.

Akses ke Google Colab dilakukan dengan masuk ke colab.research.google.com, dan file notebook dapat disimpan di Google Drive untuk kemudahan akses dan kolaborasi.

3.4.2. Pustaka Python yang Digunakan

Berikut adalah pustaka utama yang digunakan dalam penelitian ini:

a. Pustaka untuk Pengambilan dan Pengolahan Data

1. `yfinance` → Mengambil data harga saham dari Yahoo Finance.

2. pandas → Mengelola dan memproses dataset saham.

3. numpy → Operasi matematis pada data saham.

b. Pustaka untuk K-Means Clustering

1. scikit-learn (sklearn.cluster.KMeans) → Mengimplementasikan algoritma K-Means Clustering.

2. matplotlib & seaborn → Visualisasi hasil clustering.

c. Pustaka untuk Optimasi Portofolio dengan Model Markowitz

PyPortfolioOpt → Menghitung return ekspektasi, matriks kovarians, dan bobot optimal portofolio.

d. Pustaka untuk Evaluasi Portofolio

pypfopt.performance → Menghitung Sharpe Ratio, Value at Risk (VaR), dan Maximum Drawdown.

3.4.3. Instalasi dan Import Pustaka di Google Colab

Karena Google Colab sudah menyediakan sebagian besar pustaka yang dibutuhkan, instalasi tambahan hanya diperlukan untuk pustaka tertentu seperti yfinance dan PyPortfolioOpt.

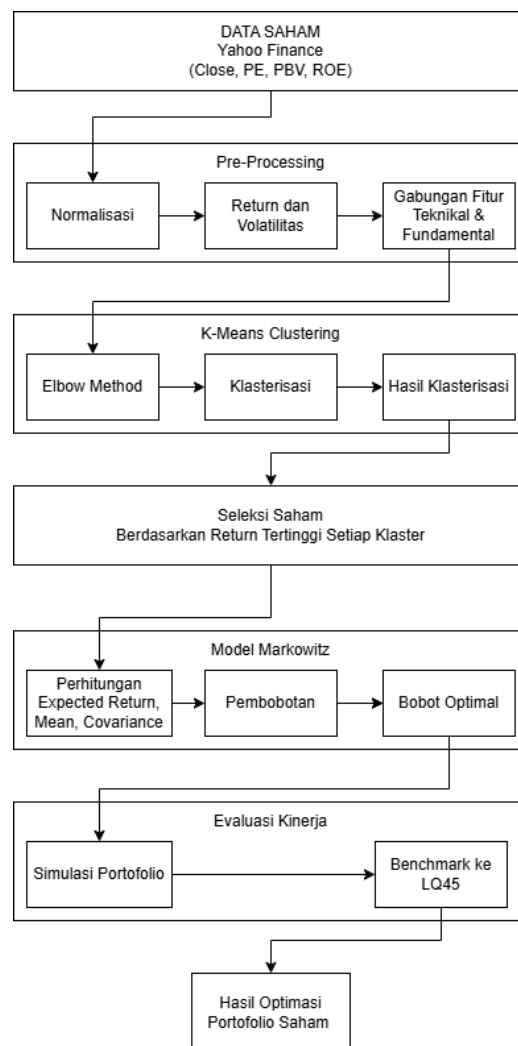
3.4.4. Alur Pemrograman di Google Colab

1. Mengambil data saham dari Yahoo Finance menggunakan yfinance.
2. Melakukan preprocessing data, termasuk pembersihan dan normalisasi.
3. Menerapkan K-Means Clustering untuk mengelompokkan saham berdasarkan fitur fundamental dan teknikal.
4. Menggunakan Model Markowitz untuk menentukan bobot optimal dalam portofolio.

5. Mengevaluasi portofolio dengan mengukur return, risiko, dan rasio performa.
6. Menampilkan hasil analisis dalam bentuk visualisasi dan tabel.

3.5. Arsitektur Penelitian

Penelitian ini menggunakan pendekatan berbasis data mining dan optimasi portofolio. Arsitektur penelitian digambarkan dalam Gambar 3.1 berikut, yang menunjukkan alur utama mulai dari pengumpulan data hingga evaluasi hasil portofolio.



Gambar 3.1. General Architecture

Adapun penjelasan pada gambar 3.1. sebagai berikut:

1. Data Saham (Yahoo Finance)

Data diperoleh dari Yahoo Finance dengan cakupan harga penutupan (Close), serta rasio keuangan seperti Price to Earnings (P/E), Price to Book Value (PBV), dan Return on Equity (ROE). Data tersebut mencerminkan aspek teknikal dan fundamental dari saham-saham yang terdaftar di Bursa Efek Indonesia.

2. Pre-Processing

Pada tahap ini, dilakukan normalisasi terhadap fitur agar memiliki skala yang seragam, perhitungan return dan volatilitas historis, serta penggabungan fitur teknikal dan fundamental untuk membentuk dataset yang siap dianalisis.

3. K-Means Clustering

Metode Elbow digunakan untuk menentukan jumlah kluster yang optimal. Selanjutnya, saham-saham dikelompokkan berdasarkan kemiripan karakteristik menggunakan algoritma K-Means. Hasil dari proses ini berupa data klusterisasi yang mengelompokkan saham ke dalam kategori homogen.

4. Seleksi Saham

Dari setiap kluster yang terbentuk, dipilih satu saham dengan nilai return historis tertinggi. Pemilihan ini bertujuan agar setiap kluster terwakili oleh saham terbaik untuk tahap optimasi portofolio.

5. Model Markowitz

Portofolio dibentuk menggunakan pendekatan Mean-Variance Optimization. Perhitungan dilakukan terhadap expected return, variance, dan covariance antar saham terpilih. Hasilnya adalah bobot investasi optimal untuk masing-masing saham dalam portofolio.

6. Evaluasi Kinerja

Evaluasi dilakukan melalui simulasi backtest pada data tahun 2024. Hasil performa portofolio dibandingkan dengan benchmark Indeks LQ45 guna mengukur keunggulan strategi optimasi yang diterapkan.

7. Hasil Optimasi Portofolio Saham

Tahap akhir dari penelitian ini adalah portofolio saham teroptimasi yang dapat digunakan sebagai strategi investasi berbasis data mining dan model Markowitz.

3.6. Waktu Penelitian

Penelitian ini dirancang untuk diselesaikan dalam satu bulan dengan menerapkan K-Means Clustering untuk pengelompokan saham dan Model Markowitz untuk optimasi portofolio. Evaluasi dilakukan dengan membandingkan portofolio dengan indeks pasar yang ada yaitu LQ45 untuk menilai kinerja portofolio. Hasil penelitian dianalisis dan disusun dalam laporan akhir guna memberikan rekomendasi optimasi portofolio berbasis data.

BAB IV

HASIL DAN PEMBAHASAN

4.1. Deskripsi Data

Penelitian ini menggunakan data sekunder berupa data harga saham harian dan rasio keuangan fundamental yang diperoleh dari sumber terbuka. Data harga saham dan data fundamental seperti PE, PBV, ROE diperoleh melalui Yahoo Finance menggunakan pustaka `yfinance` dalam bahasa pemrograman Python,

Periode data yang digunakan dalam penelitian ini adalah 1 Januari 2019 hingga 31 Desember 2024. Rentang waktu ini dipilih untuk memastikan ketersediaan data yang cukup panjang dalam mengukur karakteristik return dan risiko jangka menengah dari masing-masing saham. Dengan periode ini pula, analisis performa historis dan fundamental saham dapat dilakukan secara lebih representatif.

Jenis data yang digunakan dalam penelitian ini terdiri atas dua kategori utama, yaitu data teknikal dan data fundamental.

1. Data Teknikal:

- a. Harga penutupan harian (close price) untuk seluruh saham LQ45
- b. Return harian dan rata-rata return tahunan (mean return)
- c. Volatilitas tahunan (standar deviasi dari return harian dikalikan $\sqrt{252}$).

2. Data Fundamental:

- a. Price-to-Earnings Ratio (P/E): menggambarkan valuasi saham terhadap laba perusahaan.

- b. Price-to-Book Value (PBV): menggambarkan nilai pasar terhadap nilai buku perusahaan.
- c. Return on Equity (ROE): menunjukkan efisiensi penggunaan modal sendiri dalam menghasilkan laba.

Setelah seluruh data dikumpulkan, dilakukan proses preprocessing untuk memastikan kelayakan analisis. Langkah-langkah preprocessing yang dilakukan antara lain:

1. Menghapus saham dengan data harga atau fundamental yang tidak lengkap.
2. Normalisasi fitur numerik menggunakan teknik StandardScaler dari pustaka scikit-learn, agar semua variabel berada dalam skala yang setara.
3. Menggabungkan data teknikal dan fundamental menjadi satu DataFrame (features_all.csv) sebagai basis data untuk proses clustering.

Visualisasi dan dokumentasi tiap tahapan pengolahan data disajikan pada Gambar 4.1 hingga Gambar 4.5, yang menunjukkan proses mulai dari pengambilan harga saham, perhitungan return dan volatilitas, hingga penggabungan data teknikal dan fundamental. Proses ini menjadi fondasi utama sebelum dilakukan proses clustering saham menggunakan algoritma K-Means pada tahapan selanjutnya.

```

1 | pip install yfinance --quiet
2 | import yfinance as yf
3 | import pandas as pd
4 |
5 | # Daftar saham LQ45
6 | tickers = [
7 |     'ACES', 'ADRO', 'AKRA', 'AMRT', 'ARTO', 'ASII',
8 |     'BBCA', 'BBNI', 'BBRI', 'BBTN', 'BMRI', 'BRIS',
9 |     'BRPT', 'BSDE', 'CPIN', 'EMTK', 'ERAA', 'EXCL',
10 |    'GOTO', 'HRUM', 'ICBP', 'INCO', 'INDF', 'INKP',
11 |    'ISAT', 'ITMG', 'JFPA', 'KLBF', 'MDKA', 'MEDC',
12 |    'MIKA', 'PGAS', 'PTBA', 'PTPP', 'SCMA', 'SMGR',
13 |    'TIGO', 'TINS', 'TKIM', 'TLKM', 'TOWR', 'UNTR',
14 |    'UNVR', 'WKA'
15 | ]
16 | tickers = [kode + ".JK" for kode in tickers]
17 |
18 | # Download data
19 | start = "2019-01-01"
20 | end = "2024-12-31"
21 |
22 | data = yf.download(tickers, start=start, end=end, group_by='ticker', auto_adjust=True)
23 |
24 | # Ambil harga close
25 | close_prices = pd.DataFrame({ticker.replace(".JK", ""): data[ticker]['Close'] for ticker
26 | in tickers if ticker in data.columns.levels[0]})
27 | close_prices.dropna(axis=1, how='any', inplace=True) # hanya gunakan saham dengan data
28 | lengkap
29 | close_prices.to_csv('price_data.csv')
30 | close_prices.head()

```

Gambar 4.1. Pengambilan Harga Penutupan Saham

Pengambilan data harga saham historis pada penelitian ini dilakukan dengan menggunakan bahasa pemrograman Python, yang didukung oleh pustaka `yfinance` dan `pandas`. Pustaka `yfinance` dimanfaatkan untuk memperoleh data harga saham dari sumber daring Yahoo Finance, sementara `pandas` digunakan untuk manipulasi dan penyusunan data dalam bentuk `DataFrame`. Daftar saham yang digunakan merupakan saham-saham yang termasuk dalam indeks LQ45 di Bursa Efek Indonesia (IDX). Daftar ticker saham ini kemudian disesuaikan dengan format yang diperlukan oleh Yahoo Finance, yaitu dengan menambahkan akhiran `.JK` pada setiap kode saham.

Penelitian ini menetapkan periode pengambilan data mulai dari 1 Januari 2019 hingga 31 Desember 2024. Proses pengunduhan data dilakukan menggunakan fungsi `yf.download()` dengan parameter `group_by='ticker'`, yang memastikan data diperoleh dalam format multi-indeks untuk setiap saham. Dari keseluruhan data yang diperoleh, fokus utama penelitian ini adalah pada harga penutupan (`Close`), karena harga penutupan digunakan dalam analisis perhitungan return dan volatilitas saham. Selanjutnya, dilakukan pembersihan dan penyusunan data ke dalam `DataFrame` yang hanya memuat data harga penutupan, serta penghapusan akhiran `.JK` pada nama ticker agar lebih ringkas dan seragam.

Untuk memastikan kualitas data yang akan digunakan dalam tahap analisis lebih lanjut, dilakukan penghapusan kolom-kolom yang memiliki data kosong (`missing values`) dengan menggunakan fungsi `dropna()`. Hal ini penting agar hanya data saham yang lengkap yang dilibatkan dalam proses analisis, sehingga meningkatkan keandalan hasil penelitian. Data harga penutupan yang telah bersih dan terintegrasi kemudian disimpan dalam file CSV bernama `price_data.csv`, yang

akan menjadi dasar dalam tahapan preprocessing, pengelompokan saham menggunakan K-Means Clustering, dan optimasi portofolio dengan Model Markowitz.

Sebagai verifikasi awal, data yang telah disusun ditampilkan menggunakan fungsi `head()` untuk memastikan bahwa struktur dan format data telah sesuai. Dengan demikian, proses ini menghasilkan dataset yang telah siap dan valid untuk digunakan dalam analisis lebih lanjut. Langkah-langkah ini menjadi tahap awal yang krusial dalam penelitian, karena kualitas dan kelengkapan data historis akan mempengaruhi akurasi hasil analisis dan efektivitas portofolio yang dioptimalkan.

```

1 import pandas as pd
2 import numpy as np
3
4 # Load data harga saham yang sudah disimpan sebelumnya
5 price_data = pd.read_csv('price_data.csv', index_col=0, parse_dates=True)
6
7 # 1. Hitung return harian dari data harga penutupan
8 daily_returns = price_data.pct_change().dropna()
9
10 # 2. Hitung rata-rata return tahunan
11 mean_return = daily_returns.mean() * 252 # diasumsikan 252 hari trading dalam setahun
12
13 # 3. Hitung volatilitas tahunan (standar deviasi return harian *  $\sqrt{252}$ )
14 volatility = daily_returns.std() * np.sqrt(252)
15
16 # 4. Gabungkan return dan volatilitas ke dalam satu tabel
17 features_teknikal = pd.DataFrame({
18     'Mean_Return': mean_return,
19     'Volatility': volatility
20 })
21
22 # Simpan hasilnya
23 features_teknikal.to_csv('features_teknikal.csv')
24 features_teknikal.head()
25

```

Gambar 4.2. Perhitungan Return dan Volatilitas Saham

Dalam tahap ini, dilakukan proses perhitungan return harian, rata-rata return tahunan, serta volatilitas tahunan dari data harga saham yang telah diperoleh sebelumnya. Proses ini dilaksanakan menggunakan pustaka Python pandas dan numpy untuk memudahkan pengolahan data numerik dan tabular.

Pertama, data harga saham yang telah tersimpan dalam file CSV `price_data.csv` dibaca kembali ke dalam DataFrame menggunakan fungsi `pd.read_csv()`. File ini telah memuat data harga penutupan saham yang lengkap dan terstruktur rapi.

Langkah berikutnya adalah menghitung return harian dari data harga penutupan yang tersedia. Perhitungan ini dilakukan dengan menggunakan metode `pct_change()` dari `pandas`, yang secara otomatis menghitung persentase perubahan harga dari satu hari ke hari berikutnya. Data hasil perhitungan kemudian dibersihkan dari nilai kosong menggunakan fungsi `dropna()` untuk memastikan data yang digunakan dalam tahap analisis bersih dan lengkap.

Setelah memperoleh return harian, dilakukan perhitungan rata-rata return tahunan dengan mengalikan rata-rata return harian dengan 252, yaitu jumlah hari perdagangan aktif dalam satu tahun. Nilai ini diasumsikan sebagai standar jumlah hari trading dalam setahun di bursa saham.

Selanjutnya, volatilitas tahunan dihitung sebagai standar deviasi dari return harian yang kemudian dikalikan dengan akar kuadrat dari 252. Hal ini sesuai dengan pendekatan standar dalam perhitungan volatilitas tahunan (annualized volatility).

Hasil perhitungan rata-rata return tahunan dan volatilitas tahunan ini kemudian digabungkan ke dalam sebuah `DataFrame` baru untuk mempermudah analisis dan visualisasi lebih lanjut. `DataFrame` ini memuat dua kolom utama, yaitu 'Mean_Return' yang menunjukkan rata-rata return tahunan, dan 'Volatility' yang menunjukkan volatilitas tahunan dari masing-masing saham.

Akhirnya, data yang telah diperoleh disimpan ke dalam file CSV baru bernama `features_teknikal.csv`, yang selanjutnya dapat digunakan dalam tahap analisis clustering. Sebagai langkah verifikasi, lima baris pertama data ditampilkan menggunakan fungsi `head()`.

Tahapan ini menjadi dasar dalam mengidentifikasi karakteristik teknikal setiap saham, yang selanjutnya akan digunakan dalam tahap pengelompokan menggunakan algoritma K-Means Clustering. Proses yang sistematis ini bertujuan untuk memastikan bahwa setiap saham dinilai berdasarkan pola return dan volatilitas yang konsisten dan relevan dengan kebutuhan optimasi portofolio.

```

1 |pip install yfinance --quiet
2 |import yfinance as yf
3 |import pandas as pd
4
5 # Daftar kode saham LQ45
6 lq45_tickers = [
7     'ACES', 'ADRO', 'AKRA', 'AMRT', 'ARTO', 'ASII',
8     'BBCA', 'BBNI', 'BBRI', 'BBTN', 'BMRI', 'BRIS',
9     'BRPT', 'BSDE', 'CPIN', 'EMTK', 'ERAA', 'EXCL',
10    'GOTO', 'HRUM', 'ICBP', 'INDO', 'INDF', 'INKP',
11    'ISAT', 'ITMG', 'JPFA', 'KLBF', 'MKA', 'MEDC',
12    'MIKA', 'PGAS', 'PTBA', 'PTPP', 'SCMA', 'SMGR',
13    'TBIG', 'TINS', 'TKIM', 'TLKM', 'TOWR', 'UNTR',
14    'UNVR', 'WIKA'
15 ]
16
17 # Proses ambil data
18 results = []
19 for kode in lq45_tickers:
20     ticker = f'{kode}.JK'
21     try:
22         info = yf.Ticker(ticker).info
23         pe = info.get('trailingPE', '')
24         pbv = info.get('priceToBook', '')
25         roe = info.get('returnOnEquity', '')
26         if roe != '' and roe is not None:
27             roe *= 100 # ubah ke persen
28         results.append({
29             'Kode': kode,
30             'PE': round(pe, 2) if isinstance(pe, (float, int)) else '',
31             'PBV': round(pbv, 2) if isinstance(pbv, (float, int)) else '',
32             'ROE': round(roe, 2) if isinstance(roe, (float, int)) else ''
33         })
34     except Exception as e:
35         results.append({'Kode': kode, 'PE': '', 'PBV': '', 'ROE': ''})
36         print(f'Gagal ambil data untuk {kode}: {e}')
37
38 # Simpan ke DataFrame
39 df_fundamental = pd.DataFrame(results)
40
41 # Tampilkan dan simpan ke CSV
42 print(df_fundamental)
43 df_fundamental.to_csv('fundamental_lq45.csv', index=False)
44

```

Gambar 4.3. Pembuatan Data Fundamental

Dalam tahap ini, dilakukan proses pengambilan data fundamental saham-saham LQ45 dari Yahoo Finance menggunakan pustaka Python yfinance. Proses ini bertujuan untuk mendapatkan data rasio keuangan yang menjadi indikator fundamental penting dalam analisis saham, yaitu Price to Earnings (P/E) Ratio, Price to Book Value (PBV), dan Return on Equity (ROE).

Tahap pertama dimulai dengan mendefinisikan daftar kode saham LQ45, yang akan menjadi fokus dalam pengambilan data. Kode saham kemudian dilengkapi dengan akhiran .JK agar sesuai dengan format ticker di Yahoo Finance.

Selanjutnya, dilakukan inisialisasi list kosong results yang akan digunakan untuk menyimpan data fundamental setiap saham.

Proses pengambilan data dilakukan dengan iterasi pada setiap kode saham di dalam list lq45_tickers. Untuk setiap ticker, digunakan objek yf.Ticker untuk mengakses informasi fundamental. Kemudian, data P/E Ratio (trailingPE), PBV (priceToBook), dan ROE (returnOnEquity) diambil menggunakan metode info.get(). Nilai ROE dikalikan dengan 100 agar dinyatakan dalam persentase.

Setelah data diperoleh, nilai-nilai rasio ini diperiksa untuk memastikan bahwa nilai yang disimpan adalah bertipe numerik (float atau integer). Jika data valid, nilai akan dibulatkan hingga dua angka di belakang koma. Data fundamental kemudian disusun ke dalam dictionary yang berisi kode saham, P/E, PBV, dan ROE, lalu ditambahkan ke list results.

Jika terjadi kesalahan dalam pengambilan data, misalnya karena data tidak tersedia atau koneksi internet bermasalah, proses ini akan tetap berlanjut ke saham berikutnya, dan data yang tidak berhasil diambil akan ditandai sebagai kosong. Hal ini dilakukan dengan memanfaatkan blok try-except.

Data yang telah berhasil dikumpulkan kemudian diubah menjadi DataFrame menggunakan pustaka pandas, dengan kolom Kode, PE, PBV, dan ROE untuk mempermudah tahap analisis berikutnya.

Akhirnya, data fundamental ini disimpan ke dalam file CSV bernama fundamental_lq45.csv menggunakan fungsi to_csv() dengan parameter

`index=False` untuk menghindari penulisan indeks tambahan. Langkah ini memastikan bahwa data fundamental telah tersimpan rapi dan dapat digunakan untuk integrasi dengan data teknikal pada tahap analisis selanjutnya.

Dengan demikian, tahap ini berhasil menghasilkan dataset fundamental saham LQ45 yang bersih dan siap digunakan dalam proses analisis lebih lanjut, seperti penggabungan fitur, clustering, dan optimasi portofolio. Proses ini menjadi dasar dalam mengintegrasikan data fundamental dan teknikal untuk menghasilkan portofolio investasi yang optimal.

```
1 # Load data fundamental saham LQ45
2 fundamental_df = pd.read_csv("fundamental_lq45.csv")
3
4 # Pastikan kolom 'Kode' menjadi index agar bisa digabung dengan data teknikal
5 fundamental_df.set_index('Kode', inplace=True)
6
7 # Gabungkan fitur teknikal dan fundamental berdasarkan kode saham
8 features_all = features_teknikal.join(fundamental_df, how='inner')
9
10 # Simpan hasil gabungan
11 features_all.to_csv("features_all.csv")
12 features_all.head()
```

Gambar 4.4. Penggabungan Fitur Teknikal dan Fundamental

Pada tahap ini, dilakukan penggabungan data fundamental dan data teknikal saham LQ45 untuk membentuk dataset yang lengkap. Penggabungan ini bertujuan untuk memperoleh dataset final yang memuat seluruh fitur yang relevan, baik dari sisi teknikal (mean return dan volatilitas) maupun fundamental (P/E, PBV, dan ROE). Dataset final ini akan digunakan pada tahap analisis clustering dan optimasi portofolio.

Langkah pertama dalam penggabungan ini adalah memuat data fundamental yang telah disimpan sebelumnya dalam file `fundamental_lq45.csv` menggunakan fungsi `pd.read_csv()` dari pustaka `pandas`.

Agar data fundamental dapat digabungkan dengan data teknikal yang telah tersedia pada DataFrame `features_teknikal`, kolom Kode pada data fundamental diatur sebagai indeks. Langkah ini penting karena penggabungan akan dilakukan berdasarkan kode saham sebagai kunci utama (key) yang menjadi identitas masing-masing data.

Selanjutnya, dilakukan penggabungan antara DataFrame `features_teknikal` dan `fundamental_df` menggunakan fungsi `join()` dengan parameter `how='inner'`. Parameter ini memastikan bahwa hanya data saham yang memiliki data lengkap pada kedua tabel yang akan digabungkan, sehingga kualitas dataset final tetap terjaga.

Hasil penggabungan ini kemudian disimpan dalam file CSV bernama `features_all.csv` menggunakan fungsi `to_csv()`. File ini akan menjadi dasar dalam tahap pengelompokan saham dengan K-Means Clustering dan optimasi portofolio menggunakan Model Markowitz.

Sebagai langkah verifikasi awal, lima baris pertama dari dataset final ditampilkan menggunakan fungsi `head()` untuk memastikan integritas dan kelengkapan data.

Dengan demikian, tahapan ini berhasil menghasilkan dataset yang terintegrasi dan siap digunakan dalam tahap analisis lebih lanjut. Penggabungan data fundamental dan teknikal ini merupakan bagian penting dalam memastikan bahwa portofolio yang dibentuk tidak hanya mempertimbangkan karakteristik pergerakan harga saham (teknikal), tetapi juga kondisi fundamental perusahaan yang mendasarinya.

Berikut ini data yang tersimpan di `features_all.csv` yaitu data gabungan antara teknikal dan fundamental:

Tabel 4.1. Data Gabungan Teknikal dan Fundamental

Kode Saham	Mean Return	Volatility	PE	PBV	ROE
AKRA	0.182242707	0.409548399	10.52	2.02	16.74
AMRT	0.275736221	0.385975371	27.6	5.3	19.28
ARTO	1.137707383	0.843826652	207.34	3.12	1.52
BBCA	0.156482646	0.242376488	18.86	4.04	21.71
BBNI	0.082176807	0.333058526	7.29	0.96	13.46
BBRI	0.133330196	0.317814138	9.37	1.78	18.96
...	...	...	...	...	...
MEDC	0.275244088	0.538357523	4.27	12470.59	17.44
SCMA	0.008523434	0.48989574	21.19	1.69	5.76
SMGR	-0.111959679	0.402344607	24.29	0.4	1.61
TBIG	0.293141438	0.412531407	32.76	4.46	12.41
UNVR	-0.177713385	0.334581793	21.36	31.06	121.82

Cuplikan isi `features_all.csv` yang menampilkan nilai mean return, volatilitas, P/E, PBV, dan ROE untuk masing-masing saham LQ45 setelah penggabungan data.

4.2. Pengelompokan Saham dengan K-Means Clustering

4.2.1. Normalisasi dan Seleksi Fitur

Setelah data teknikal dan fundamental digabungkan ke dalam satu DataFrame (`features_all.csv`), langkah pertama yang dilakukan adalah normalisasi. Hal ini penting karena setiap fitur memiliki skala yang berbeda—misalnya, volatilitas bisa berada pada skala 0–1, sementara ROE bisa bernilai puluhan. Normalisasi dilakukan menggunakan metode Standard Scaler, yang mengubah setiap nilai menjadi distribusi dengan mean 0 dan standar deviasi 1.

Fitur yang digunakan dalam proses clustering adalah:

1. Mean Return (rata-rata return tahunan)
2. Volatility (standar deviasi return tahunan)
3. Price to Earnings Ratio (P/E)

4. Price to Book Value (PBV)
5. Return on Equity (ROE)

Pemilihan fitur ini bertujuan untuk mencerminkan kombinasi karakteristik risiko, return, dan valuasi saham, sehingga kelompok saham yang terbentuk memiliki kemiripan secara fundamental maupun teknikal.

```
1 # Hapus baris (saham) yang memiliki nilai NaN
2 features_all_clean = features_all.dropna()
3
4 # Lanjutkan normalisasi
5 from sklearn.preprocessing import StandardScaler
6 from sklearn.cluster import KMeans
7 import matplotlib.pyplot as plt
8
9 scaler = StandardScaler()
10 features_scaled = scaler.fit_transform(features_all_clean)
```

Gambar 4.5. Preprocessing Data untuk Clustering

Pada tahap ini, dilakukan pembersihan dan normalisasi data yang telah digabungkan sebelumnya. Data yang akan digunakan dalam proses clustering harus dalam kondisi bersih dan terstandarisasi agar hasil pengelompokan menjadi valid dan dapat diinterpretasikan dengan baik.

Langkah pertama adalah menghapus baris (saham) yang memiliki nilai kosong (NaN) pada salah satu fitur. Hal ini penting untuk memastikan bahwa setiap saham yang akan dianalisis memiliki data yang lengkap pada seluruh fitur fundamental dan teknikal. Proses ini dilakukan menggunakan fungsi `dropna()` dari pustaka `pandas`.

Setelah memastikan bahwa dataset bersih dari nilai kosong, tahap berikutnya adalah melakukan normalisasi data menggunakan metode standardisasi. Normalisasi ini bertujuan untuk menyamakan skala dari seluruh fitur yang digunakan dalam analisis clustering, sehingga setiap fitur memiliki kontribusi yang

seimbang. Proses normalisasi dilakukan menggunakan pustaka scikit-learn dengan objek `StandardScaler`.

Pada tahap ini, objek `StandardScaler` digunakan untuk menghitung rata-rata dan standar deviasi dari setiap fitur, kemudian mentransformasikan nilai-nilai pada dataset agar memiliki distribusi normal dengan rata-rata nol dan standar deviasi satu. Data yang telah dinormalisasi ini selanjutnya akan digunakan dalam penerapan algoritma K-Means Clustering untuk pengelompokan saham berdasarkan karakteristik fundamental dan teknikal yang serupa.

Dengan demikian, tahap pembersihan dan normalisasi data ini menjadi dasar penting untuk menghasilkan analisis clustering yang akurat dan sesuai dengan tujuan penelitian.

4.2.2. Penentuan Jumlah Kluster

```

1 # Elbow Method untuk tentukan jumlah kluster
2 inertia = []
3 K_range = range(1, 11)
4 for k in K_range:
5     kmeans = KMeans(n_clusters=k, random_state=42, n_init=10)
6     kmeans.fit(features_scaled)
7     inertia.append(kmeans.inertia_)
8
9 # Visualisasi Elbow Curve
10 plt.figure(figsize=(8, 5))
11 plt.plot(K_range, inertia, marker='o')
12 plt.title('Elbow Method for Optimal K')
13 plt.xlabel('Number of Clusters (K)')
14 plt.ylabel('Inertia')
15 plt.grid(True)
16 plt.show()

```

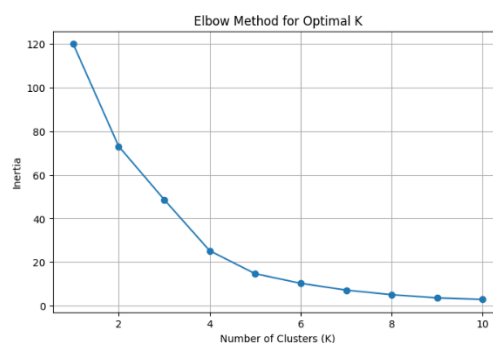
Gambar 4.6. Elbow Method untuk Menentukan Jumlah Kluster

Setelah data saham dibersihkan dan dinormalisasi, langkah selanjutnya adalah menentukan jumlah kluster yang optimal sebelum penerapan algoritma K-Means Clustering. Penentuan jumlah kluster ini dilakukan dengan menggunakan metode Elbow Method, yang bertujuan untuk menemukan titik optimal (*elbow*) di mana penambahan jumlah kluster selanjutnya tidak lagi memberikan pengurangan yang signifikan pada nilai inertia.

Pada tahap ini, digunakan variasi jumlah klaster dari 1 hingga 10. Untuk setiap nilai k , dilakukan penerapan K-Means Clustering dengan parameter $\text{random_state}=42$ dan $\text{n_init}=10$ untuk memastikan hasil yang konsisten dan akurat. Nilai inertia dari setiap percobaan (yang menunjukkan total jarak antar data ke centroid klaster terdekat) disimpan dalam list inertia untuk dianalisis.

Setelah memperoleh nilai inertia untuk setiap jumlah klaster, dilakukan visualisasi hasilnya dalam bentuk grafik Elbow Curve menggunakan pustaka `matplotlib`. Visualisasi ini mempermudah dalam mengidentifikasi jumlah klaster optimal dengan melihat titik di mana penurunan nilai inertia mulai melambat.

Grafik Elbow Curve yang dihasilkan akan menunjukkan titik elbow, yaitu jumlah klaster optimal yang akan digunakan dalam tahap clustering selanjutnya. Penentuan nilai k ini menjadi tahap penting karena akan memengaruhi kualitas pengelompokan saham dan efektivitas diversifikasi portofolio yang dihasilkan.



Gambar 4.7. Grafik Elbow Method

Grafik Elbow menunjukkan adanya tekukan signifikan pada $K = 4$, yang menunjukkan bahwa empat klaster merupakan jumlah yang paling efisien untuk memisahkan saham ke dalam kelompok dengan karakteristik yang serupa.

4.2.3. Hasil Clustering dan Karakteristik Kluster

Setelah jumlah kluster ditetapkan, algoritma K-Means Clustering dijalankan dengan K=4. Saham-saham LQ45 kemudian dikelompokkan ke dalam empat kluster berbeda berdasarkan kemiripan karakteristik.

```
1 from sklearn.cluster import KMeans
2
3 # Jalankan K-Means dengan K = 4
4 kmeans = KMeans(n_clusters=4, random_state=42, n_init=10)
5 clusters = kmeans.fit_predict(features_scaled)
6
7 # Tambahkan label kluster ke DataFrame asli
8 features_all_clean = features_all.dropna()
9 features_all_clean['Cluster'] = clusters
10
11 # Simpan hasil clustering
12 features_all_clean.to_csv("clustered_stocks.csv")
13
14 # Tampilkan hasil clustering
15 features_all_clean.sort_values('Cluster').head(10)
16
```

Gambar 4.8. Kode Program Klustering

Pada tahap ini dilakukan penerapan algoritma K-Means Clustering dengan parameter `n_clusters=4`, `random_state=42`, dan `n_init=10` untuk memastikan konsistensi hasil pengelompokan.

Hasil pengelompokan atau label kluster yang diperoleh dari algoritma K-Means disimpan ke dalam variabel `clusters`. Label kluster ini kemudian ditambahkan sebagai kolom baru bernama 'Cluster' ke dalam DataFrame asli `features_all_clean`. Langkah ini bertujuan untuk mengintegrasikan hasil pengelompokan dengan data teknikal dan fundamental saham yang telah dibersihkan dan dinormalisasi.

Selanjutnya, hasil akhir pengelompokan disimpan ke dalam file CSV bernama `clustered_stocks.csv` untuk memudahkan proses analisis dan interpretasi lebih lanjut.

Sebagai langkah verifikasi, ditampilkan beberapa baris awal data saham yang telah dilengkapi dengan label klaster, diurutkan berdasarkan kolom 'Cluster'. Visualisasi ini membantu dalam memahami komposisi saham pada masing-masing klaster.

Dengan demikian, tahap ini berhasil menghasilkan data saham yang telah dikelompokkan ke dalam empat klaster berdasarkan karakteristik teknikal dan fundamentalnya. Hasil clustering ini akan menjadi dasar dalam tahap seleksi saham terbaik dari masing-masing klaster yang akan digunakan dalam proses optimasi portofolio pada tahap selanjutnya.

Dari Program yang dijalankan diatas, didapati hasil clustering yang disimpan di file clustered_stocks.csv, diantara lain:

Tabel 4.2. Hasil Klasterisasi

Kode Saham	Mean Return	Volatility	PE	PBV	ROE	Cluster
ARTO	1.137707	0.843827	207.34	3.12	1.52	0
AKRA	0.182243	0.409548	10.52	2.02	16.74	1
AMRT	0.275736	0.385975	27.6	5.3	19.28	1
BBCA	0.156483	0.242376	18.86	4.04	21.71	1
BBNI	0.082177	0.333059	7.29	0.96	13.46	1
...	...	...	...	...	...	...
MEDC	0.275244	0.538358	4.27	12470.59	17.44	1
TBIG	0.293141	0.412531	32.76	4.46	12.41	1
SMGR	-0.11196	0.402345	24.29	0.4	1.61	1
BRPT	0.276716	0.576774	74.26	41666.67	2.96	2
UNVR	-0.177713	0.334582	21.36	31.06	121.8	3

Dari hasil klastering, didapati karakteristik setiap klaster berdasarkan data fundamental dan teknikal, berikut adalah interpretasi gaya investasi dari masing-masing klaster:

Tabel 4.3. Interpretasi Gaya Investasi Setiap Kluster

Klaster	Gaya Investasi	Karakteristik	Contoh Saham
0	High Risk, High Return	Return tinggi, volatilitas tinggi	ARTO
1	Balanced Growth	Return dan risiko moderat	ISAT
2	Moderate-Aggressive	Return tinggi, risiko sedang	BRPT
3	Defensive / Konservatif	Return rendah, risiko rendah	UNVR

Dari hasil ini, terlihat bahwa pendekatan clustering berhasil membagi saham ke dalam kelompok yang secara alami mencerminkan strategi investasi yang berbeda, sehingga sangat berguna untuk proses seleksi saham secara lebih terstruktur pada tahap selanjutnya.

4.3. Seleksi Saham dari Setiap Klaster

Setelah seluruh saham dalam indeks LQ45 berhasil dikelompokkan ke dalam empat klaster menggunakan algoritma K-Means, langkah berikutnya adalah memilih satu saham terbaik dari setiap klaster. Tujuan dari pemilihan ini adalah untuk memperoleh portofolio yang terdiversifikasi secara representatif, dengan masing-masing saham mencerminkan karakteristik khas dari klaster yang terbentuk.

4.3.1. Kriteria Seleksi

Kriteria yang digunakan dalam seleksi saham dari setiap klaster adalah nilai rata-rata return tahunan (Mean Return). Saham dengan nilai Mean Return tertinggi dalam masing-masing klaster dipilih sebagai wakil dari klaster tersebut. Strategi ini digunakan untuk memastikan bahwa saham yang dipilih memiliki potensi keuntungan yang paling tinggi di antara rekan-rekannya dalam klaster yang sama.

4.3.2. Implementasi Seleksi Saham Otomatis

Proses seleksi dilakukan secara otomatis menggunakan skrip Python. Berikut adalah potongan kode yang digunakan:

```

1 import pandas as pd
2
3 # Load ulang hasil clustering yang sudah disimpan sebelumnya
4 clustered_df = pd.read_csv("clustered_stocks.csv", index_col=0)
5
6 # Lanjutkan seleksi saham terbaik per kluster
7 def pilih_saham_terbaik_per_kluster(df, return_col='Mean_Return', cluster_col='Cluster'):
8     best_per_cluster = df.groupby(cluster_col)[return_col].idxmax()
9     return df.loc[best_per_cluster].sort_values(cluster_col)
10
11 selected_stocks = pilih_saham_terbaik_per_kluster(clustered_df)
12 selected_stocks.to_csv("selected_stocks_per_cluster.csv")
13
14 # Tampilkan hasil seleksi
15 selected_stocks
16

```

Gambar 4.9. Kode Otomatisasi Pemilihan Terbaik Setiap Klaster

Pada tahap ini, dilakukan proses seleksi saham terbaik dari setiap klaster yang telah terbentuk. Tujuan dari tahap ini adalah untuk memilih satu saham unggulan dari masing-masing klaster berdasarkan kinerja return-nya, sehingga diperoleh portofolio yang lebih representatif dan berpotensi menghasilkan return yang optimal.

Langkah pertama adalah memuat ulang data hasil clustering yang telah disimpan dalam file `clustered_stocks.csv`. Data ini dimuat ke dalam DataFrame `clustered_df` menggunakan pustaka `pandas`.

Untuk melakukan seleksi saham terbaik dari setiap klaster, dibuat sebuah fungsi `pilih_saham_terbaik_per_kluster()` yang menerima parameter DataFrame (`df`), kolom return (`return_col`), dan kolom klaster (`cluster_col`). Fungsi ini bekerja dengan cara mencari indeks saham dengan nilai return rata-rata (`Mean_Return`) tertinggi pada masing-masing klaster. Langkah ini menggunakan metode `groupby()` untuk mengelompokkan data berdasarkan klaster, kemudian mengambil saham dengan `idxmax()` yang menunjukkan nilai return maksimum.

Setelah fungsi dibuat, dilakukan pemanggilan fungsi tersebut dengan menggunakan data `clustered_df`. Hasilnya adalah DataFrame `selected_stocks` yang

berisi saham-saham terbaik dari masing-masing klaster, yang menjadi kandidat utama dalam portofolio yang akan dioptimasi pada tahap selanjutnya.

Hasil seleksi kemudian disimpan dalam file CSV `selected_stocks_per_cluster.csv` untuk dokumentasi dan digunakan sebagai data input pada tahap optimasi portofolio. Sebagai langkah verifikasi, hasil seleksi saham terbaik ditampilkan menggunakan perintah `selected_stocks` untuk memastikan data telah sesuai.

Dengan demikian, tahap ini menghasilkan daftar saham unggulan yang mewakili setiap klaster. Proses seleksi ini sangat penting karena akan menjadi dasar dalam pembentukan portofolio yang terdiversifikasi dengan optimal, serta mendukung pencapaian keseimbangan antara risiko dan return investasi.

4.3.3. Hasil Seleksi Saham

Berdasarkan hasil clustering dan proses seleksi otomatis, berikut adalah empat saham terpilih:

Tabel 4.4. Hasil Seleksi Saham

Klaster	Saham	Mean Return (Tahun)	Volatilitas	Karakteristik
0	ARTO	Sangat Tinggi	Tinggi	Agresif
1	ISAT	Sedang	Sedang	Seimbang
2	BRPT	Tinggi	Sedang	Moderat
3	UNVR	Negatif	Rendah	Defensif

Namun, hasil optimasi portofolio yang dilakukan pada tahap selanjutnya menunjukkan bahwa saham UNVR tidak mendapatkan bobot investasi (0%), yang menandakan bahwa saham tersebut tidak berkontribusi positif dalam peningkatan Sharpe Ratio portofolio. Oleh karena itu, saham ini akan dieliminasi dari tahap optimasi meskipun sempat terpilih dalam proses seleksi awal.

4.4. Optimasi Portofolio dengan Model Markowitz

Setelah memperoleh empat saham terpilih hasil seleksi dari masing-masing klaster, langkah selanjutnya adalah melakukan optimasi portofolio dengan menggunakan Model Markowitz. Model ini dikenal sebagai pendekatan Mean-Variance Optimization (MVO), yang bertujuan untuk menentukan komposisi bobot investasi yang memaksimalkan return dengan risiko terkendali, atau dalam konteks penelitian ini: memaksimalkan Sharpe Ratio.

4.4.1. Persiapan Data untuk Optimasi

Data harga saham historis dari saham-saham terpilih dikumpulkan kembali untuk periode Januari 2019 hingga Desember 2024, sama seperti rentang data pelatihan sebelumnya. Saham yang digunakan dalam proses ini adalah:

1. ARTO (Klaster 0)
2. BRPT (Klaster 2)
3. ISAT (Klaster 1)
4. UNVR (Klaster 3)

Namun, setelah proses optimasi dilakukan, diketahui bahwa saham UNVR memperoleh bobot nol dalam portofolio optimal, sehingga tidak dimasukkan dalam simulasi backtest di tahap berikutnya.

```

1 import yfinance as yf
2 import numpy as np
3 import pandas as pd
4 from scipy.optimize import minimize
5 import matplotlib.pyplot as plt
6
7 # Baca saham hasil seleksi dari tiap klaster
8 selected_df = pd.read_csv("selected_stocks_per_cluster.csv", index_col=0)
9 symbols = selected_df.index.tolist()
10 symbols_yf = [kode + ".JK" for kode in symbols]
11
12 # Unduh harga penutupan historis
13 data_raw = yf.download(symbols_yf, start="2019-01-01", end="2024-12-31", auto_adjust=True)
14
15 price_data = data_raw['Close'] if 'Close' in data_raw else data_raw
16 price_data.dropna(inplace=True)
17
18 # Hitung return harian, rata-rata tahunan, dan kovarian tahunan
19 returns = price_data.pct_change().dropna()
20 mean_returns = returns.mean() * 252
21 cov_matrix = returns.cov() * 252
22 num_assets = len(symbols)

```

Gambar 4.10. Pengambilan Data dan Perhitungan Return

Pada tahap ini, dilakukan persiapan data untuk optimasi portofolio menggunakan Model Markowitz. Saham-saham yang digunakan adalah hasil seleksi terbaik dari masing-masing klaster yang telah diperoleh sebelumnya. Langkah ini bertujuan untuk memperoleh dataset yang akurat, lengkap, dan siap digunakan dalam perhitungan bobot optimal portofolio.

Langkah pertama adalah memuat data saham hasil seleksi yang telah disimpan dalam file `selected_stocks_per_cluster.csv`. Data ini dimuat ke dalam DataFrame `selected_df`, dan kode saham diekstrak sebagai list `symbols` untuk memudahkan proses pengunduhan data harga historis selanjutnya.

Setelah itu, dilakukan pengunduhan data harga penutupan historis dari Yahoo Finance menggunakan pustaka `yfinance` dengan parameter waktu mulai dari 1 Januari 2019 hingga 31 Desember 2024. Data harga penutupan ini akan digunakan sebagai dasar dalam menghitung return dan matriks kovarian antar saham.

Langkah selanjutnya adalah menghitung return harian, rata-rata return tahunan, dan kovarian tahunan dari data harga saham. Return harian dihitung sebagai persentase perubahan harga harian menggunakan fungsi `pct_change()`. Nilai rata-rata return tahunan diperoleh dengan mengalikan rata-rata return harian

dengan 252 (jumlah hari perdagangan aktif dalam setahun). Demikian pula, matriks kovarian return tahunan dihitung dari kovarian return harian yang dikalikan dengan 252.

Akhirnya, jumlah saham yang akan digunakan dalam portofolio dicatat menggunakan variabel `num_assets`. Langkah ini penting sebagai dasar dalam penyusunan fungsi objektif dan pembatasan (constraint) pada proses optimasi portofolio yang akan dilakukan pada tahap berikutnya.

Dengan demikian, tahap ini menghasilkan dataset teknikal yang lengkap, yaitu rata-rata return tahunan dan matriks kovarian antar saham hasil seleksi. Dataset ini menjadi input utama dalam perhitungan bobot optimal portofolio berdasarkan pendekatan Mean-Variance Optimization (MVO) yang akan dilakukan pada tahap optimasi berikutnya.

4.4.2. Persiapan Data untuk Optimasi

Model Markowitz digunakan dengan tujuan memaksimalkan Sharpe Ratio, yaitu rasio antara return ekspektasi dikurangi risk-free rate terhadap volatilitas portofolio. Optimasi dilakukan menggunakan metode Sequential Least Squares Programming (SLSQP).

```

1 #Fungsi objektif: negatif Sharpe Ratio
2 def negative_sharpe(weights, mean_returns, cov_matrix, risk_free_rate=0.03):
3     portfolio_return = np.dot(weights, mean_returns)
4     portfolio_volatility = np.sqrt(np.dot(weights.T, np.dot(cov_matrix, weights)))
5     sharpe_ratio = (portfolio_return - risk_free_rate) / portfolio_volatility
6     return -sharpe_ratio # Negatif untuk minimisasi
7
8 # Constraint dan batasan
9 constraints = ({'type': 'eq', 'fun': lambda x: np.sum(x) - 1})
10 bounds = tuple((0, 1) for _ in range(num_assets))
11 init_guess = num_assets * [1. / num_assets]
12
13 # Jalankan optimasi
14 opt_result = minimize(negative_sharpe, init_guess, args=(mean_returns, cov_matrix),
15                       method='SLSQP', bounds=bounds, constraints=constraints)

```

Gambar 4.11. Persiapan Data

Pada tahap ini, dilakukan proses optimasi portofolio menggunakan pendekatan Mean-Variance Optimization (MVO) berdasarkan Model Markowitz. Tujuan optimasi ini adalah untuk menentukan bobot optimal setiap aset dalam portofolio sehingga diperoleh rasio Sharpe tertinggi, yang menjadi indikator portofolio dengan return risiko yang paling efisien.

Langkah pertama adalah mendefinisikan fungsi objektif yang digunakan dalam proses optimasi, yaitu negatif Sharpe Ratio. Fungsi ini menerima parameter bobot (weights), rata-rata return tahunan (mean_returns), matriks kovarian tahunan (cov_matrix), dan tingkat suku bunga bebas risiko (risk_free_rate). Dalam implementasinya, return portofolio dihitung sebagai perkalian antara bobot portofolio dan rata-rata return tahunan. Volatilitas portofolio dihitung sebagai akar dari hasil perkalian matriks bobot dengan matriks kovarian. Rasio Sharpe kemudian dihitung dengan cara membagi selisih return portofolio terhadap tingkat suku bunga bebas risiko dengan volatilitas portofolio. Fungsi ini mengembalikan nilai negatif rasio Sharpe untuk keperluan minimisasi.

Selanjutnya, ditetapkan constraint dan batasan (boundary) dalam proses optimasi. Constraint yang digunakan adalah bahwa total bobot portofolio harus berjumlah satu ($\text{sum}(\text{weights}) = 1$), yang memastikan portofolio sepenuhnya diinvestasikan. Batasan bobot untuk setiap aset diatur antara 0 dan 1, yang berarti tidak diperbolehkan adanya short selling dalam portofolio.

Proses optimasi dijalankan menggunakan fungsi `minimize()` dari pustaka `scipy.optimize` dengan metode 'SLSQP'. Fungsi ini meminimasi fungsi objektif negatif Sharpe Ratio dengan mempertimbangkan constraint dan batasan yang telah

ditetapkan sebelumnya. Sebagai titik awal optimasi (`init_guess`), digunakan bobot yang sama rata untuk semua saham di portofolio.

Hasil dari optimasi ini adalah bobot portofolio optimal (`opt_result.x`) yang memaksimalkan rasio Sharpe. Langkah ini menjadi puncak dalam perhitungan portofolio optimal, yang mempertimbangkan keseimbangan antara potensi return dan tingkat risiko dari seluruh saham yang terpilih. Dengan demikian, penelitian ini menghasilkan portofolio yang tidak hanya terdiversifikasi, tetapi juga disusun secara optimal untuk mencapai efisiensi risiko-return yang lebih tinggi.

4.4.3. Hasil Bobot Portofolio

```

1 # Simpan dan tampilkan bobot
2 optimal_weights = opt_result.x
3 optimal_portfolio = pd.Series(optimal_weights, index=symbols)
4 optimal_portfolio.to_csv("optimal_portfolio_weights.csv")
5 print("=== Bobot Optimal Portofolio ===")
6 print(optimal_portfolio)
7
8 # Visualisasi pie chart komposisi
9 optimal_portfolio.plot(kind='pie', autopct='%1.1f%%', startangle=90, figsize=(6, 6))
10 plt.title('Komposisi Optimal Portofolio Markowitz')
11 plt.ylabel('')
12 plt.tight_layout()
13 plt.show()

```

Gambar 4.12. Program Visualisasi

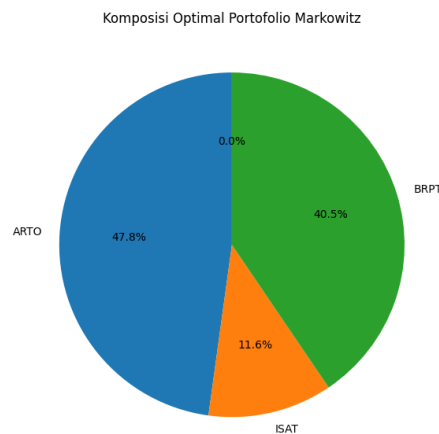
Setelah proses optimasi portofolio selesai dilakukan, tahap selanjutnya adalah menyimpan dan memvisualisasikan bobot optimal dari masing-masing saham dalam portofolio. Bobot optimal tersebut diperoleh dari hasil keluaran fungsi optimasi yang disimpan dalam variabel `opt_result.x`.

Langkah pertama adalah menyusun bobot optimal ke dalam bentuk Series dari pustaka pandas, dengan indeks berupa simbol saham hasil seleksi. Data bobot ini kemudian disimpan ke dalam file CSV bernama `optimal_portfolio_weights.csv` agar dapat digunakan untuk dokumentasi atau analisis lanjutan.

Untuk mempermudah interpretasi visual terhadap distribusi aset dalam portofolio, dilakukan visualisasi dalam bentuk diagram pie. Diagram ini memperlihatkan komposisi alokasi dana pada masing-masing saham berdasarkan bobot optimal hasil perhitungan Model Markowitz.

Visualisasi ini memberikan gambaran proporsi investasi yang disarankan pada masing-masing saham, berdasarkan prinsip efisiensi risiko-return. Melalui representasi ini, investor dapat memahami dengan jelas distribusi aset yang membentuk portofolio optimal, sekaligus sebagai bahan pertimbangan strategis dalam implementasi keputusan investasi.

Hasil dari proses optimasi menunjukkan bobot optimal portofolio sebagai berikut:



Gambar 4.13. Hasil Optimasi Markowitz

Berikut adalah bobot investasi dari masing-masing saham hasil optimasi:

Tabel 4.5. Bobot Investasi

Saham	Bobot Optimal	Karakteristik
ARTO	47.8%	High return, high risk (agresif)
BRPT	40.5%	Moderat, stabil
ISAT	11.6%	Menengah, seimbang
UNVR	0.00%	Tidak dipilih (return negatif)

Hasil ini menunjukkan bahwa portofolio mengutamakan saham-saham dengan return tinggi meskipun memiliki risiko yang lebih besar. Saham UNVR tidak dimasukkan karena menurunkan nilai Sharpe Ratio secara keseluruhan.

4.5. Simulasi Kinerja Portofolio Tahun 2024

Setelah portofolio optimal berhasil dibentuk melalui proses optimasi menggunakan Model Markowitz, langkah berikutnya adalah mengukur efektivitas portofolio tersebut dalam kondisi pasar yang sesungguhnya. Untuk itu, dilakukan simulasi backtest terhadap kinerja portofolio selama periode 1 Januari 2024 hingga 31 Desember 2024. Tujuan dari simulasi ini adalah untuk mengevaluasi sejauh mana komposisi portofolio optimal dapat memberikan return yang kompetitif di tengah dinamika pasar aktual.

4.5.1. Metode Simulasi Backtest

Simulasi dilakukan dengan cara menerapkan bobot hasil optimasi terhadap data harga aktual saham terpilih pada tahun 2024. Saham yang digunakan dalam simulasi terdiri dari tiga saham dengan bobot tidak 0, yaitu:

1. ARTO (47.82%)
2. BRPT (40.52%)
3. ISAT (11.64%)

Saham UNVR tidak dimasukkan karena memperoleh bobot 0% dalam proses optimasi. Nilai awal portofolio diasumsikan sebesar 100, dan dihitung pertumbuhan nilainya dari hari ke hari berdasarkan return harian dari masing-masing saham yang dikalikan dengan bobotnya.

Berikut adalah kode program Python yang digunakan untuk proses simulasi backtest:

```

1 import yfinance as yf
2 import pandas as pd
3 import numpy as np
4 import matplotlib.pyplot as plt
5
6 # === 1. Masukkan bobot portofolio hasil optimasi ===
7 portfolio_weights = {
8     'ARTO': 0.491817,
9     'ISAT': 0.126222,
10    'BRPT': 0.381962,
11 }
12
13 tickers = [kode + ".JK" for kode in portfolio_weights.keys()]
14 start_date = "2024-01-01"
15 end_date = "2024-12-31"
16 data = yf.download(tickers, start=start_date, end=end_date, auto_adjust=True)['Close']
17 data.dropna(inplace=True)
18
19 # === 2. Hitung return portofolio harian dan akumulatif ===
20 returns = data.pct_change().dropna()
21 weights_array = np.array(list(portfolio_weights.values()))
22 portfolio_returns = returns.dot(weights_array)
23 portfolio_value = (1 + portfolio_returns).cumprod() * 100
24
25 # === 3. Visualisasi hasil backtest ===
26 plt.figure(figsize=(10, 6))
27 plt.plot(portfolio_value, label='Portofolio Markowitz')
28 plt.title('Simulasi Backtest Portofolio Tahun 2024')
29 plt.xlabel('Tanggal')
30 plt.ylabel('Nilai Portofolio (Index Base 100)')
31 plt.grid(True)
32 plt.legend()
33 plt.tight_layout()
34 plt.show()
35
36 # === 4. Tampilkan performa akhir ===
37 total_return = (portfolio_value[-1] / portfolio_value[0] - 1) * 100
38 print(f"Total Return Portofolio 2024: {total_return:.2f}%")
39

```

Gambar 4.14. Simulasi Backtest Portofolio Tahun 2024

Untuk mengevaluasi kinerja portofolio yang telah dioptimasi, dilakukan simulasi backtest dengan menggunakan data harga penutupan saham pada tahun 2024. Tahap ini bertujuan untuk menilai sejauh mana portofolio yang disusun melalui optimasi Model Markowitz mampu memberikan return yang optimal pada periode waktu yang ditentukan.

Langkah awal adalah mendefinisikan bobot portofolio yang diperoleh dari hasil optimasi. Bobot tersebut ditetapkan secara manual pada dictionary `portfolio_weights`, yang berisi proporsi alokasi dana untuk masing-masing saham.

Selanjutnya, dilakukan pengunduhan data harga penutupan historis untuk saham-saham tersebut selama periode 1 Januari 2024 hingga 31 Desember 2024 menggunakan pustaka `yfinance`. Data yang telah diperoleh dibersihkan dari nilai kosong agar data yang digunakan dalam simulasi lengkap dan valid.

Setelah data harga saham diperoleh, dilakukan perhitungan return harian portofolio menggunakan pendekatan `linear weighted return`. Bobot portofolio dikonversi ke dalam bentuk array, kemudian digunakan untuk menghitung return harian portofolio melalui perkalian dot product antara return harian setiap saham dan bobot masing-masing saham. Nilai portofolio kumulatif dihitung dengan cara mengalikan faktor pertumbuhan harian secara akumulatif (`cumprod()`), dikalikan 100 agar disajikan dalam bentuk indeks.

Untuk memberikan pemahaman visual terhadap kinerja portofolio, hasil backtest divisualisasikan dalam bentuk grafik line plot. Grafik ini menunjukkan perubahan nilai portofolio setiap hari sepanjang tahun 2024, dengan sumbu X sebagai waktu (tanggal) dan sumbu Y sebagai nilai portofolio dalam bentuk indeks (basis 100).

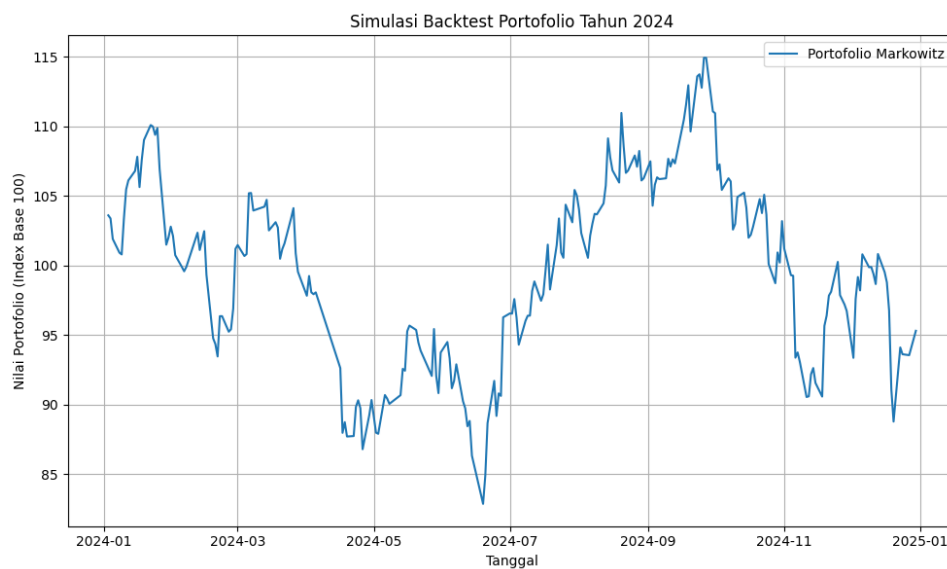
Akhirnya, untuk memperoleh ringkasan kinerja, dihitung total return portofolio sepanjang periode simulasi. Nilai return kumulatif dihitung dari selisih nilai portofolio akhir dengan awal, kemudian dinyatakan dalam bentuk persentase.

Dengan demikian, tahap ini berhasil memberikan gambaran kinerja portofolio hasil optimasi pada periode yang telah ditentukan, sekaligus menjadi evaluasi praktis terhadap efektivitas Model Markowitz yang diterapkan dalam penelitian ini.

4.5.2. Hasil dan Interpretasi

Hasil simulasi ditampilkan dalam Gambar 4.15, yang memperlihatkan fluktuasi nilai portofolio selama tahun 2024. Dari grafik tersebut terlihat bahwa nilai portofolio sempat meningkat di pertengahan tahun, namun mengalami penurunan tajam menjelang akhir tahun.

Berdasarkan hasil perhitungan, portofolio mengalami penurunan nilai sebesar 8.02% sepanjang tahun 2024. Penurunan ini mencerminkan kondisi pasar yang kurang stabil selama periode tersebut, namun perlu diperhatikan bahwa kinerja portofolio ini masih lebih baik dibandingkan indeks LQ45, yang akan dibahas pada subbab berikutnya.



Gambar 4.15. Hasil Simulasi

Simulasi ini menunjukkan bahwa meskipun strategi portofolio tidak menghasilkan return positif, penggunaan metode optimasi portofolio berhasil menekan tingkat kerugian dibandingkan pasar secara umum. Selain itu, simulasi ini membuktikan bahwa strategi Markowitz yang dikombinasikan dengan seleksi

saham berbasis clustering memiliki potensi sebagai metode yang lebih disiplin dan terukur dalam manajemen investasi.

4.6. Perbandingan Kinerja Portofolio dengan Indeks LQ45

Untuk mengetahui sejauh mana efektivitas portofolio hasil optimasi dengan Model Markowitz, dilakukan perbandingan kinerja dengan Indeks LQ45 sebagai benchmark. Indeks LQ45 dipilih karena mewakili 45 saham paling likuid dan berkapitalisasi besar di Bursa Efek Indonesia, sehingga cocok dijadikan tolok ukur umum untuk kinerja investasi di pasar saham Indonesia.

4.6.1. Metode Perbandingan

Data harga penutupan indeks LQ45 diunduh dari Yahoo Finance menggunakan kode simbol `^JKLQ45` untuk periode yang sama, yaitu 1 Januari hingga 31 Desember 2024. Nilai indeks tersebut kemudian dinormalisasi ke angka dasar 100 agar dapat dibandingkan secara proporsional dengan portofolio.

```

1 # === Tambahkan data indeks LQ45 ===
2 benchmark = yf.download("^JKLQ45", start=start_date, end=end_date, auto_adjust=True)
3 benchmark.dropna(inplace=True)
4
5 # Normalisasi ke index 100
6 benchmark_value = (benchmark / benchmark.iloc[0]) * 100
7
8 # === Gabungkan dan visualisasikan ===
9 plt.figure(figsize=(10, 6))
10 plt.plot(portfolio_value, label='Portofolio Markowitz')
11 plt.plot(benchmark_value, label='Indeks LQ45', linestyle='--')
12 plt.title('Perbandingan Kinerja: Portofolio vs Indeks LQ45 (2024)')
13 plt.xlabel('Tanggal')
14 plt.ylabel('Nilai (Index Base 100)')
15 plt.legend()
16 plt.grid(True)
17 plt.tight_layout()
18 plt.show()
19
20 # Hitung return total benchmark
21 benchmark_return = (benchmark_value.iloc[-1] / benchmark_value.iloc[0] - 1) * 100
22 print(f"Total Return LQ45 2024: {benchmark_return.item():.2f}%")
23

```

Gambar 4.16. Program Perbandingan Kinerja Portofolio dan Indeks LQ45

Untuk memberikan konteks yang lebih komprehensif terhadap performa portofolio hasil optimasi, dilakukan perbandingan kinerja portofolio dengan indeks

acuan pasar, yaitu Indeks LQ45. Tujuan dari tahap ini adalah untuk mengevaluasi apakah portofolio yang dibentuk mampu mengungguli performa pasar.

Langkah awal adalah pengambilan data harga penutupan harian indeks LQ45 sepanjang tahun 2024 menggunakan pustaka `yfinance`. Data yang diperoleh kemudian dibersihkan dari nilai kosong untuk memastikan keandalan data yang digunakan.

Nilai indeks LQ45 kemudian dinormalisasi agar memiliki basis 100 pada awal periode, memudahkan perbandingan kinerja dengan portofolio yang juga dinormalisasi pada basis yang sama.

Selanjutnya, dilakukan visualisasi performa portofolio dan indeks LQ45 dalam bentuk grafik. Grafik ini memperlihatkan pergerakan nilai portofolio dan indeks LQ45 sepanjang periode 2024, memudahkan identifikasi tren dan keunggulan portofolio relatif terhadap pasar.

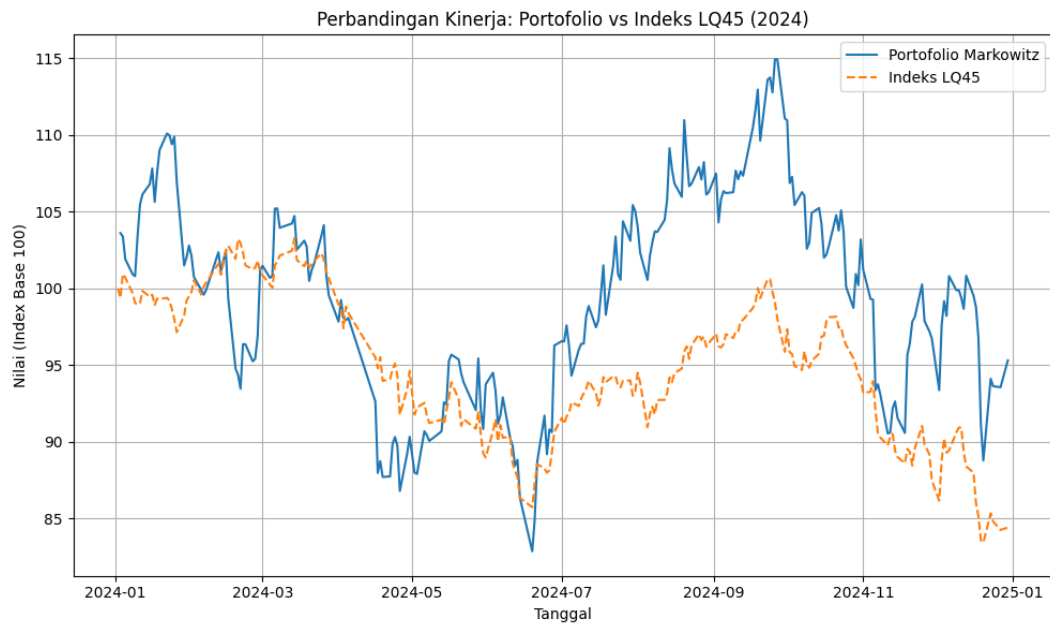
Sebagai evaluasi akhir, dilakukan perhitungan total return Indeks LQ45 sepanjang periode tersebut. Nilai return ini dihitung sebagai persentase kenaikan nilai indeks dari awal hingga akhir tahun.

Dengan demikian, penelitian ini memberikan analisis komprehensif mengenai keunggulan portofolio hasil optimasi Model Markowitz dibandingkan dengan benchmark pasar. Hasil ini menjadi indikator penting dalam menilai efektivitas portofolio yang dihasilkan dalam konteks investasi riil di pasar modal.

4.6.2. Hasil Perbandingan

Hasil dari perbandingan ditampilkan pada Gambar 4.17, yang memperlihatkan grafik pertumbuhan nilai portofolio hasil optimasi dan indeks LQ45 selama tahun 2024. Dari grafik tersebut, terlihat bahwa meskipun keduanya

mengalami fluktuasi dan tren penurunan, portofolio Markowitz berhasil menunjukkan kinerja yang lebih baik secara konsisten dibanding indeks LQ45.



Gambar 4.17. Perbandingan Kinerja

Berdasarkan perhitungan Total Return Portofolio Markowitz yang dijalankan dengan backtest pada data 2024 mengalami penurunan nilai 8.02%, sedangkan Total Return Indeks LQ45 dengan pengujian yang sama mengalami penurunan nilai hingga 15.60%.

Penurunan yang lebih kecil pada portofolio yang telah di rancang menggunakan markowitz menunjukkan bahwa alokasi bobot yang diperoleh lebih efisien dalam mengelola risiko dibanding hanya mengikuti indeks pasar. Hal ini menunjukkan potensi dari pendekatan berbasis data dan optimasi dalam strategi investasi serta pemanfaatan teknologi informasi dalam meningkatkan ekonomi.

BAB V

KESIMPULAN DAN SARAN

5.1. Kesimpulan

Penelitian ini bertujuan untuk menguji efektivitas kombinasi metode K-Means Clustering dan Model Markowitz dalam mengoptimasi portofolio saham di Bursa Efek Indonesia. Berdasarkan hasil analisis dan implementasi program yang telah dilakukan, maka dapat disimpulkan beberapa hal sebagai berikut:

1. Data historis harga saham dan data fundamental dari saham-saham yang tergabung dalam indeks LQ45 berhasil digunakan sebagai dasar untuk membentuk fitur teknikal (mean return dan volatilitas) serta fitur fundamental (P/E, PBV, dan ROE). Periode data yang digunakan mencakup Januari 2019 hingga Desember 2024.
2. Proses clustering menggunakan algoritma K-Means berhasil mengelompokkan saham ke dalam empat kluster optimal, yang masing-masing menunjukkan gaya investasi yang berbeda, seperti agresif, defensif, dan seimbang. Seleksi saham dilakukan secara otomatis dengan memilih saham dengan return tertinggi di setiap kluster, menghasilkan empat saham: ARTO, BRPT, ISAT, dan UNVR.
3. Proses optimasi portofolio menggunakan Model Markowitz menunjukkan bahwa bobot optimal yang diberikan kepada saham-saham tersebut memprioritaskan saham dengan return tinggi, seperti ARTO dan BRPT. Saham UNVR memperoleh bobot 0%, sehingga tidak dimasukkan dalam portofolio akhir.

4. Simulasi kinerja portofolio untuk tahun 2024 menunjukkan bahwa strategi portofolio hasil optimasi mendapatkan penurunan nilai hingga 8.02%, sementara indeks LQ45 mengalami penurunan yang lebih besar yakni sebesar 15.60%. Hal ini menunjukkan bahwa strategi yang dibangun mampu menekan risiko kerugian lebih baik daripada pasar secara umum, sehingga dapat dikatakan bahwa kombinasi K-Means dan Model Markowitz efektif dalam meningkatkan efisiensi portofolio.
5. Secara keseluruhan, penelitian ini membuktikan bahwa pendekatan berbasis data mining dan teori portofolio modern dapat diterapkan secara praktis untuk membangun strategi investasi yang lebih terukur dan defensif terhadap volatilitas pasar, khususnya di pasar saham Indonesia.

5.2. Saran

Berdasarkan hasil penelitian dan pengalaman implementasi yang diperoleh, penulis memberikan beberapa saran sebagai berikut:

1. Periode pengujian dapat diperluas ke beberapa tahun mendatang agar strategi dapat diuji secara lebih baik terhadap perubahan pasar.
2. Bagi investor individu dan pengelola dana, metode yang digunakan dalam penelitian ini dapat dijadikan sebagai acuan dalam membangun portofolio yang seimbang antara risiko dan return, terutama di tengah ketidakpastian pasar.
3. Penelitian ini juga dapat dikembangkan lebih lanjut untuk membandingkan performa strategi ini dengan metode optimasi lainnya

4. Diharapkan penelitian ini dapat menjadi rangka awal penerapan otomatisasi strategi portofolio berbasis Python dan data publik, yang dapat dikembangkan ke sistem rekomendasi atau aplikasi keuangan.

DAFTAR PUSTAKA

- Abrar, A. S., Kunaifi, A., & SE, M. (2021). *Diversifikasi portofolio saham syariah indonesia menggunakan algoritma k-means clustering: studi kasus pandemi covid-19* (Skripsi Sarjana, Institut Teknologi Sepuluh Nopember).
- Ady, S. U., Susilowati, S., & Farida, I. (2022). Penyuluhan pengenalan analisa fundamental pada keputusan investasi saham. *Transformasi: Jurnal Pengabdian Pada Masyarakat*, 2(1), 18-31.
<https://doi.org/10.31764/transformasi.v2i1.8099>
- Ariyatma, R. D., & Fahmi, S. (2023). Data mining menggunakan multiple regression untuk prediksi harga saham netflix. *Jurnal Saintekom: Sains, Teknologi, Komputer dan Manajemen*, 13(2), 184-192.
<https://doi.org/10.33020/saintekom.v13i2.419>
- Azizah, K. N., Saepudin, D., & Gunawan, P. H. (2021). Optimasi portofolio saham LQ45 dengan mempertimbangkan prediksi return menggunakan metode holt-winter. *E-Proceeding of Engineering*, 8(5), 10776–10784.
- Gubu, L., Rosadi, D., & Abdurakhman, A. (2021). Pembentukan portofolio saham menggunakan klastering time series k-medoid dengan ukuran jarak dynamic time warping. *Jurnal Aplikasi Statistika & Komputasi Statistik*, 13(2), 35-46. <https://doi.org/10.34123/jurnalasks.v13i2.295>
- Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., ... & Oliphant, T. E. (2020). Array programming with NumPy. *Nature*, 585(7825), 357-362.

- Inaku, R. F., & Chandra, J. C. (2023). Implementasi data mining dalam prediksi harga saham menggunakan metode long short-term memory (lstm). *Jurnal Ticom: Technology of Information and Communication*, 12(1), 1-7. <https://doi.org/10.70309/ticom.v12i1.99>
- Jum'an (2024). Indikator fundamental dan teknikal sebagai dasar pengambilan keputusan dalam berinvestasi saham. *Jurnal Ekonomi STIEP*, 9(2), 47-56. <https://doi.org/10.54526/jes.v9i2.348>
- Markowitz, H. (1952). Portfolio Selection. *The Journal of Finance*, 7(1), 77–91.
- McKinney, W. (2017). *Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython* (2nd ed.). O'Reilly Media.
- Nuha, S. U., Cahyadi, N., & Purnomo, S. W. (2023). Pengenalan langkah awal berinvestasi dalam pasar modal di era milenial. *Jurnal Pengabdian Masyarakat dan Lingkungan*, 2(1), 21-24. <http://dx.doi.org/10.30587/jpml.v2i1.6109>
- Nuraini, N. P., Simatupang, W., & Dasman, S. (2024). Analisis risiko investasi saham melalui diversifikasi portofolio secara domestik dan internasional. *Margin: Jurnal Lentera Manajemen Keuangan*, 2(01), 37-44. <https://doi.org/10.59422/margin.v2i01.259>
- Pratama, Y., Sulistianingsih, E., Debataraja, N. N., & Imro'ah, N. (2024). K-means clustering dan mean variance efficient portfolio dalam portofolio saham. *Jambura Journal of Probability and Statistics*, 5(1), 24–30.
- Pratama, Y., Sulistianingsih, E., Debataraja, N. N., & Imroâ, N. (2024). K-means clustering dan mean variance efficient portfolio dalam portofolio

- saham. *Jambura Journal of Probability and Statistics*, 5(1), 24-30.
<https://doi.org/10.37905/jjps.v5i1.20298>
- Riyandi, A., Aripin, A., Ardiansyah, I. N., Dany, R., & Yusrizal, Y. (2023). Analisis data mining untuk prediksi harga saham: perbandingan metode regresi linier dan pola historis. *Jurnal Teknologi Sistem Informasi*, 4(2), 278-288.
<https://doi.org/10.35957/jtsi.v4i2.5158>
- Samsudin, A., Aritonang, C., Munthe, G. R., Monalisa, W., & Hutasoit, Y. G. (2023). Analisis risiko investasi saham melalui diversifikasi portofolio secara domestik dan internasional. *El-Mal: Jurnal Kajian Ekonomi & Bisnis Islam*, 4(5), 1330-1351. <https://doi.org/10.47467/elmal.v4i5.2895>
- Sari, I. P., Al-Khowarizmi, A., & Sulaiman, O. K. (2023). Implementation of data classification using k-means algorithm in clustering stunting cases. *JCoSITTE*, 4(2), 402-412. <https://doi.org/10.30596/jcositte.v4i2.15765>
- Sari, I. P., Batubara, I. H., & Hanif, I. (2021). Cluster analysis using k-means algorithm and fuzzy c-means clustering for grouping students' abilities in online learning process. *JCoSITTE*, 2(1), 139-144.
<https://doi.org/10.30596/jcositte.v2i1.6504>
- Silitonga, A. I., Nasution, M. A., & Rizwinie, K. S. (2025). Klasterisasi penyebaran base transceiver station menggunakan k-means clustering. SEMNAS RISTEK, Jakarta. <https://doi.org/10.30998/semnasristek.v9i1.7947>
- Sukamto, A. S., Setiawan, W., & Pratama, E. E. (2023). Data mining untuk pengelompokan saham pada sektor energi dengan metode k-means. *JEPIN (Jurnal Edukasi dan Penelitian Informatika)*, 9(1), 76-81.

- Sunariyah. (2022). Pengaruh model mean-variance dalam optimasi portofolio saham. *Jurnal Ekonomi & Keuangan*, 14(2), 145–157.
- Tohendry, D., & Jollyta, D. (2023). Penerapan algoritma k-means clustering untuk pengelompokkan saham berdasarkan price earning ratio dan price to book value. *Jurnal Mahasiswa Aplikasi Teknologi Komputer dan Informasi (JMApTeKsi)*, 5(1), 1-7.
- Usman, D. R., Ramadhan, M., Hutasuhut, M., Jaya, H., Gunawan, R., & Kusnasari, S. (2024). Implementasi data mining untuk memprediksi pergerakan harga saham BRI dengan menggunakan metode regresi linier berganda. *Jurnal Teknologi Sistem Informasi dan Sistem Komputer TGD*, 7(1), 151-159.
<https://doi.org/10.53513/jsk.v7i1.9605>
- VanderPlas, J. (2016). *Python Data Science Handbook: Essential Tools for Working with Data*. O'Reilly Media.
- Wardhani, R. S., Vehtasvili, S. E., Aprilian, R. I., Yanto, S. E., Suhdi, S. S. T., Anggraeni Yunita, S. E., & Duwi Agustina, S. E. (2022). *Mengenal Saham*. Penerbit K-Media.