

Data-Driven Portfolio Optimization using K-Means and Markowitz Model: Evidence from LQ45 Stocks

Mutiara Akbar Nasution¹, Al-Khowarizmi²

^{1,2}Program Studi Teknologi Informasi, Universitas Muhammadiyah Sumatera Utara, Indonesia
Email: mutiaraakbarst03@gmail.com; alkhowarizmi@umsu.ac.id

ABSTRAK

Penelitian ini bertujuan untuk mengoptimalkan portofolio saham dengan pendekatan berbasis data melalui integrasi algoritma K-Means Clustering dan Model Markowitz. Data yang digunakan mencakup indikator fundamental dan teknikal dari saham-saham dalam indeks LQ45 periode 2019 hingga 2024. Proses dimulai dengan normalisasi dan ekstraksi fitur, diikuti dengan klusterisasi saham menggunakan algoritma K-Means. Dari empat kluster yang terbentuk, masing-masing dipilih satu saham dengan rata-rata return tertinggi. Setelah seleksi saham, bobot portofolio kemudian dioptimasi menggunakan Model Markowitz dengan pendekatan *mean-variance* tanpa *short selling*. Hasil optimasi menunjukkan alokasi terbesar pada saham ARTO, BRPT, dan ISAT. Evaluasi kinerja dilakukan melalui simulasi *backtest* sepanjang tahun 2024, yang menunjukkan portofolio hanya mengalami penurunan 8,02%, lebih kecil dibandingkan penurunan indeks LQ45 sebesar 15,60%. Hasil ini menegaskan bahwa integrasi pendekatan data mining dan optimasi kuantitatif mampu menghasilkan strategi portofolio yang lebih adaptif dan defensif terhadap tekanan pasar.

Keyword: K-Means Clustering; Model Markowitz; Optimasi Portofolio; Diversifikasi; Bursa Efek Indonesia

ABSTRACT

This study aims to optimize stock portfolio allocation through a data-driven approach by integrating the K-Means Clustering algorithm and the Markowitz Model. The dataset includes technical and fundamental indicators of LQ45 index stocks from 2019 to 2024. The process begins with data normalization and feature extraction, followed by stock clustering using the K-Means algorithm. From the four resulting clusters, the top-performing stock with the highest average return is selected from each. Portfolio weights are then optimized using the Markowitz Model under a mean-variance framework without short selling. The optimization results allocate the largest weights to ARTO, BRPT, and ISAT. Performance evaluation through a backtest simulation in 2024 shows that the portfolio experienced only an 8.02% decline, outperforming the LQ45 index which dropped by 15.60%. These findings underscore the potential of integrating data mining and quantitative optimization methods to improve diversification efficiency and strengthen portfolio resilience during market downturns.

Keyword: K-Means Clustering; Markowitz Model; Portfolio Optimization; Diversification; Indonesia Stock Exchange

Corresponding Author:

Mutiara Akbar Nasution,
Universitas Muhammadiyah Sumatera Utara,
Jl. Kapten Muchtar Basri No.3, Glugur Darat II, Kec. Medan Tim., Kota
Medan, Sumatera Utara 20238, Indonesia
Email: mutiaraakbarst03@gmail.com



1. INTRODUCTION

Investasi saham telah menjadi pilihan populer bagi investor yang ingin memperoleh keuntungan jangka panjang. Namun, dinamika pasar yang kompleks dan dipengaruhi oleh berbagai faktor ekonomi, teknikal, dan fundamental membuat pengambilan keputusan investasi memerlukan pendekatan yang sistematis dan berbasis data (Utami et al., 2019). Salah satu strategi yang banyak diterapkan adalah optimasi portofolio saham, yang bertujuan untuk mencapai keseimbangan antara risiko dan imbal hasil investasi.

Teori portofolio yang diperkenalkan oleh Harry Markowitz pada tahun 1952, yang dikenal dengan Mean-Variance Portfolio Theory (MVP), menjadi landasan dalam pembentukan portofolio optimal. Model ini memanfaatkan pendekatan kuantitatif untuk meminimalkan risiko dengan tetap mempertahankan tingkat pengembalian tertentu melalui prinsip diversifikasi (Markowitz, 1952; Pratama et al., 2024). Namun, penerapan Model Markowitz secara konvensional memiliki kelemahan dalam tahap seleksi saham, karena kurang mempertimbangkan homogenitas karakteristik antar saham yang terpilih (Sukanto et al., 2023).

Untuk mengatasi keterbatasan tersebut, pendekatan data mining seperti *K-Means Clustering* digunakan sebagai teknik pra-pemilihan saham. Algoritma ini memungkinkan pengelompokan saham berdasarkan kesamaan karakteristik fundamental dan teknikal, seperti *Price-to-Earnings* (P/E) Ratio, *Price-to-Book Value* (PBV), *Return on Equity* (ROE), serta return dan volatilitas historis (Tohendry & Jollyta, 2023). Melalui proses clustering, diversifikasi portofolio menjadi lebih efektif karena masing-masing saham dalam portofolio berasal dari klaster yang berbeda dan memiliki korelasi yang lebih rendah.

Beberapa penelitian sebelumnya telah menunjukkan bahwa kombinasi *K-Means Clustering* dan Model Markowitz mampu menghasilkan portofolio yang lebih optimal dibandingkan pendekatan konvensional. Utami et al. (2019) dan Pratama et al. (2024) menyatakan bahwa strategi ini dapat meningkatkan stabilitas dan efisiensi portofolio. Selain itu, integrasi metode ini dengan pemrograman Python dan pustaka seperti Scikit-Learn serta PyPortfolioOpt mempermudah implementasi analisis berbasis data secara efisien dan dapat direproduksi (Sukanto et al., 2023).

Penggunaan integrasi K-Means dan Markowitz berbasis Python untuk pasar Indonesia belum banyak dibahas dalam jurnal ekonomi lokal. Penelitian ini bertujuan untuk mengombinasikan algoritma *K-Means Clustering* dan Model Markowitz dalam membentuk portofolio saham optimal di Bursa Efek Indonesia. Data yang digunakan mencakup indikator teknikal dan fundamental pada saham-saham likuid dari tahun 2019 hingga 2024. Diharapkan, pendekatan ini dapat memberikan kontribusi nyata terhadap strategi investasi berbasis data serta memperkuat pemanfaatan teknologi dalam manajemen keuangan modern.

2. RESEARCH METHOD

Penelitian ini menggunakan pendekatan kuantitatif deskriptif dengan tujuan mengembangkan strategi optimasi portofolio saham berbasis data mining dan teori portofolio modern. Data yang digunakan dalam penelitian ini bersifat sekunder dan diperoleh melalui Yahoo Finance, mencakup periode Januari 2019 hingga Desember 2024. Data yang dikumpulkan terdiri dari harga penutupan saham harian (close price) sebagai data teknikal, serta rasio keuangan fundamental seperti Price-to-Earnings (P/E) Ratio, Price-to-Book Value (PBV), dan Return on Equity (ROE). Saham yang dianalisis merupakan bagian dari indeks LQ45 yang mewakili saham-saham paling likuid di Bursa Efek Indonesia.

Penelitian ini diawali dengan proses *preprocessing* data, meliputi pembersihan nilai hilang, normalisasi data, serta perhitungan return dan volatilitas saham sebagai fitur teknikal. Selanjutnya, algoritma *K-Means Clustering* diterapkan untuk mengelompokkan saham berdasarkan kemiripan karakteristik fundamental dan teknikal. Jumlah klaster (K) ditentukan dengan menggunakan *Elbow Method*, sedangkan evaluasi kualitas pengelompokan dilakukan melalui perhitungan *Silhouette Score*.

Dari masing-masing klaster yang terbentuk, dipilih satu saham terbaik berdasarkan rata-rata return tahunan tertinggi. Seleksi ini dilakukan secara otomatis dengan pemrograman Python agar bersifat objektif dan replikatif. Saham-saham terpilih kemudian dioptimasi menggunakan Model Markowitz dengan pendekatan *Mean-Variance Optimization* untuk menentukan bobot portofolio yang optimal. Optimasi dilakukan dengan memaksimalkan rasio Sharpe menggunakan algoritma *Sequential Least Squares Programming* (SLSQP) dari pustaka *scipy.optimize*, dengan batasan tidak diperbolehkannya short selling dan total bobot portofolio sebesar 100%.

Untuk mengukur efektivitas strategi ini, dilakukan evaluasi kinerja melalui simulasi backtest menggunakan data aktual harga saham sepanjang tahun 2024. Hasil kinerja portofolio kemudian dibandingkan dengan performa indeks LQ45 sebagai benchmark, baik dari sisi return kumulatif maupun stabilitas fluktuasi. Seluruh proses analisis dilakukan dalam lingkungan Google Colab menggunakan bahasa pemrograman Python dan pustaka-pustaka seperti Pandas, NumPy, Scikit-Learn, yfinance, dan PyPortfolioOpt (McKinney, 2017; Sukanto et al., 2023; Pratama et al., 2024). Seluruh proses dilakukan secara terotomatisasi dan dapat direproduksi menggunakan Google Colab, memungkinkan pendekatan ini diadaptasi oleh investor ritel dan akademisi.

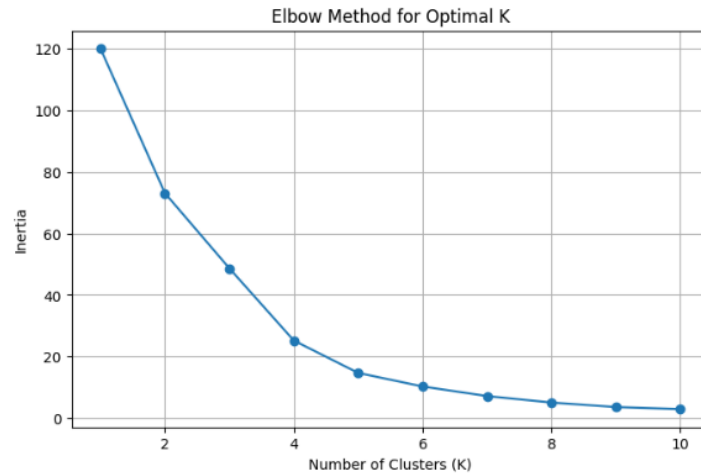
3. RESULTS AND DISCUSSION

Penelitian ini bertujuan untuk membentuk portofolio saham optimal dengan mengombinasikan algoritma *K-Means Clustering* dan Model Markowitz. Data yang digunakan mencakup saham-saham dalam

indeks LQ45, yang diproses secara bertahap mulai dari normalisasi data teknikal dan fundamental, pengelompokan saham, hingga optimasi bobot investasi.

A. Hasil Klasterisasi Saham

Dengan menggunakan metode Elbow, ditentukan jumlah kluster optimal sebanyak empat.



Gambar 1. Elbow Method

Masing-masing kluster memiliki karakteristik unik berdasarkan return, volatilitas, dan rasio keuangan. Silhouette Score mendekati angka 0,5 yang menunjukkan pengelompokan sudah cukup baik. Klasterisasi dilakukan menggunakan algoritma *K-Means* dari pustaka Scikit-Learn.

Tabel 1. Hasil K-Means Clustering

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

B. Seleksi Saham dari Tiap Kluster

Setelah kluster terbentuk, dilakukan seleksi saham terbaik dari masing-masing kluster berdasarkan return rata-rata tertinggi menggunakan skrip Python. Saham yang terpilih untuk tahap optimasi adalah:

- Kluster 0: ARTO
- Kluster 1: BRPT
- Kluster 2: ISAT
- Kluster 3: UNVR

Proses ini dilakukan otomatis menggunakan kode groupby dan idxmax dari pustaka Pandas, memastikan seleksi yang objektif dan berbasis data.

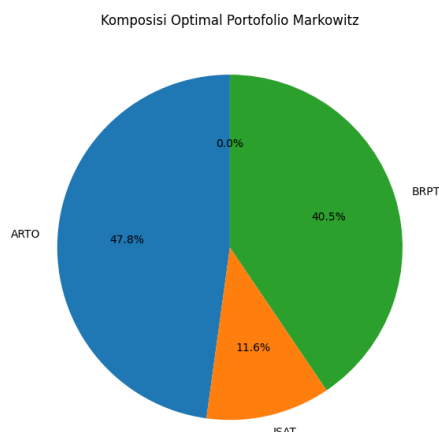
Tabel 2. Karakteristik Setiap Kluster

Kluster	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

C. Hasil Optimasi Model Markowitz

Keempat saham tersebut dioptimasi menggunakan pendekatan mean-variance untuk memaksimalkan rasio Sharpe. Proses optimasi dilakukan menggunakan algoritma SLSQP dari pustaka scipy.optimize, dengan

constraint total bobot 100% dan tanpa short selling. Hasil optimasi menunjukkan distribusi bobot sebagai berikut:

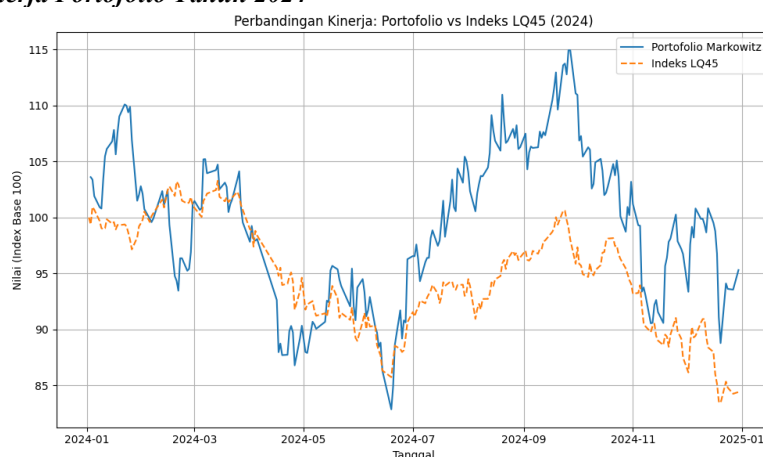


Gambar 2. Hasil Optimasi Markowitz

- **ARTO:** 47.82%
- **BRPT:** 40.52%
- **ISAT:** 11.64%
- **UNVR:** 0.00%

UNVR tidak masuk ke dalam komposisi akhir karena secara statistik tidak memberikan kontribusi optimal terhadap peningkatan rasio return terhadap risiko dalam portofolio.

D. Simulasi Kinerja Portofolio Tahun 2024



Gambar 3. Perbandingan Portofolio dengan indeks LQ45

Portofolio hasil optimasi kemudian diuji performanya dengan simulasi backtest menggunakan data harga saham aktual sepanjang tahun 2024. Hasil backtest menunjukkan bahwa portofolio mengalami penurunan nilai sebesar 8,02% selama periode tersebut. Meskipun negatif, penurunan ini masih lebih kecil dibandingkan dengan kinerja indeks LQ45 yang pada periode sama tercatat mengalami penurunan sebesar 15,60%.

Dengan demikian, meskipun pasar secara keseluruhan mengalami koreksi tajam pada tahun 2024, portofolio hasil optimasi terbukti lebih resilien terhadap tekanan pasar. Hal ini menunjukkan bahwa pendekatan diversifikasi berbasis kluster dan optimasi rasio Sharpe dapat membantu mengurangi dampak negatif dari fluktuasi pasar secara keseluruhan. Kelebihan dari strategi ini adalah pendekatannya yang tidak subjektif dan mudah diotomasi, namun keterbatasan tetap ada, misalnya ketergantungan pada data historis dan asumsi tidak adanya short selling.

Hasil ini konsisten dengan temuan Sukamto et al. (2023) dan Pratama et al. (2024) yang menyatakan bahwa penggunaan algoritma clustering dapat meningkatkan efektivitas diversifikasi dan membantu Model Markowitz menghasilkan alokasi aset yang lebih realistis dan menguntungkan. Penelitian ini juga menunjukkan bahwa integrasi metode berbasis Python dapat membantu investor dan analis keuangan dalam mengambil keputusan yang lebih terstruktur, berbasis data, dan dapat direproduksi secara efisien.

4. CONCLUSION

Penelitian ini menunjukkan bahwa integrasi algoritma K-Means Clustering dan Model Markowitz merupakan pendekatan yang efektif dalam membentuk portofolio saham yang optimal di Bursa Efek Indonesia. Melalui pengelompokan saham berdasarkan karakteristik teknikal dan fundamental, serta optimasi bobot investasi menggunakan pendekatan *mean-variance*, diperoleh alokasi aset yang lebih terdiversifikasi dan efisien. Komposisi portofolio akhir menunjukkan distribusi bobot terbesar pada saham ARTO, BRPT, dan ISAT, dengan saham UNVR dieliminasi karena kontribusinya yang tidak optimal terhadap peningkatan rasio Sharpe.

Evaluasi kinerja melalui simulasi *backtest* tahun 2024 menunjukkan bahwa portofolio hanya mengalami penurunan nilai sebesar 8,02%, dibandingkan dengan penurunan indeks LQ45 sebesar 15,60%. Hasil ini menandakan bahwa pendekatan berbasis data dan optimasi kuantitatif dapat meningkatkan resiliensi portofolio terhadap tekanan pasar yang menurun.

Secara keseluruhan, penelitian ini memberikan kontribusi terhadap pemanfaatan metode data mining dan teori portofolio modern dalam konteks pasar modal Indonesia. Integrasi antara algoritma K-Means dan Model Markowitz juga memperlihatkan potensi besar dalam membangun strategi investasi berbasis data yang sistematis, objektif, dan dapat direproduksi menggunakan perangkat analitik seperti Python. Penelitian selanjutnya dapat mengeksplorasi penggunaan teknik clustering yang lebih kompleks atau integrasi indikator makroekonomi sebagai variabel tambahan untuk meningkatkan ketepatan alokasi portofolio.

REFERENCES

- 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. <https://openlibrarypublications.telkomuniversity.ac.id/index.php/engineering/article/view/15653>
- 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, 357–362. <https://doi.org/10.1038/s41586-020-2649-2>
- 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., Debatara, 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. <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*, 9(1). <https://doi.org/10.30998/semnasristek.v9i1.7947>
- Sukanto, 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. <https://doi.org/10.26418/jp.v9i1.62509>
- 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., Vebtasvili, 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.