

Analisis Trend Saham LQ-45 selama Pandemi Covid-19 dengan K-Means Clustering

Prima Hanjasto¹✉

¹PT. Igaras

primahanjasto@gmail.com

Abstract

Stocks are one of the investment instruments that have great risks and benefits. Large profits in a very fast time often make investors forget about the risks that may occur. Dynamic price changes at any time require in-depth analysis to determine the decision to buy, hold or sell. The COVID-19 pandemic has had a major effect on economic growth, investment and culture. This study aims to group stocks that have the same trend during the pandemic so that they can get returns and minimize the risk of loss. The stock price data in this study is sourced from finance.yahoo.com with a time span from 2019 to 2021. The stocks to be analyzed are stocks belonging to the LQ-45 stock category from 2019 to 2021. will be grouped with K-means Clustering. This study groups LQ-45 stocks into 3 clusters which are the right considerations by investors in stock investments during the Covid-19 pandemic.

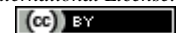
Keywords: Analysis, Stocks, LQ-45, Covid-19, K-Means Clustering.

Abstrak

Saham merupakan salah satu instrument investasi yang memiliki resiko dan keuntungan yang besar. Keuntungan yang besar dalam waktu yang sangat cepat sering membuat investor lupa tentang resiko yang mungkin terjadi. Perubahan harga yang dinamis setiap saat membutuhkan analisis mendalam untuk menentukan keputusan membeli, menahan atau menjual. Pandemi COVID-19 memberikan efek besar terhadap pertumbuhan ekonomi, investasi dan budaya. Penelitian ini bertujuan untuk mengelompokkan saham-saham yang memiliki trend sama selama pandemi sehingga bisa mendapatkan return dan meminimalisir resiko kerugian. Data harga saham dalam penelitian ini bersumber dari finance.yahoo.com dengan rentang waktu dari tahun 2019 sampai dengan 2021. Saham-saham yang akan dianalisis adalah saham-saham yang termasuk kategori saham LQ-45 dari tahun 2019 sampai dengan 2021. Saham-saham tersebut akan dikelompokkan dengan K-means Clustering. Penelitian ini mengelompokkan saham LQ-45 kedalam 3 klaster yang menjadi pertimbangan tepat oleh investor dalam investasi saham selama masa pandemi Covid-19.

Kata kunci: Analisis, Saham, LQ-45, Covid-19, K-Means Clustering.

INFEB is licensed under a Creative Commons 4.0 International License.



1. Pendahuluan

Teknologi informasi merupakan sebuah kebutuhan bukan lagi sebuah tren bagi kehidupan manusia saat ini [1]. Perangkat teknologi informasi berkembang dengan pesat yang mengakibatkan hampir semua aspek kehidupan bermasyarakat tergantung dengan teknologi informasi, seperti: smartphone, internet, aplikasi rapat online, aplikasi marketplace, aplikasi transportasi, aplikasi kesehatan, aplikasi pendidikan, aplikasi investasi dan keuangan [2].

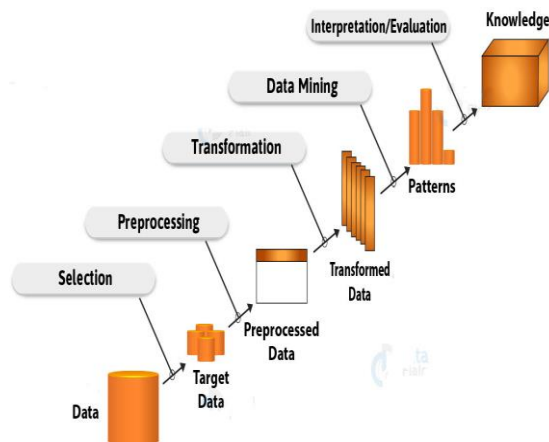
Aplikasi investasi dan keuangan online sangat berkembang yang membuat masyarakat bisa dengan mudah dan cepat dalam mengakses [3]. Saham merupakan salah satu jenis investasi yang sangat disenangi oleh masyarakat [4]. Keuntungan yang besar dalam waktu yang singkat merupakan daya tarik saham bagi investor, tetapi banyak investor maupun calon investor yang lupa bahwa saham merupakan investasi dengan resiko paling besar [5]. Hasil analisa yang

cermat terhadap arah tren harga saham dapat menjadi pertimbangan bagi investor dalam berinvestasi sehingga bisa meminimalisir resiko kerugian yang mungkin terjadi [6].

Pandemi Covid-19 yang terjadi tahun 2019 sangat berdampak besar terhadap ekonomi yang mengakibatkan ketidakstabilan dalam investasi [7]. Analisa tren harga saham dilakukan dengan melihat dan mempelajari pola harga saham yang terjadi di masa lalu berdasarkan periode tertentu [8]. Tren saham secara umum dikategorikan tiga kelompok, yaitu uptrend, notrend dan downtrend [9].

Data histori saham bisa dengan sangat mudah didapat pada saat ini, karena perkembangan teknologi big data [10]. Big data merupakan kumpulan data yang sangat banyak, tetapi tidak semua data yang bisa diolah menjadi sebuah informasi atau pengetahuan baru [11]. Knowledge Discovery in Database (KDD) merupakan serangkaian proses pencarian pengetahuan baru atau

informasi dari sebuah big data [12]. KDD merupakan proses identifikasi dan validasi pola baru yang berguna sehingga dapat dipahami dari database [13]. Teknik pencarian pengetahuan baru atau informasi dalam database disebut *Data Mining* [14].



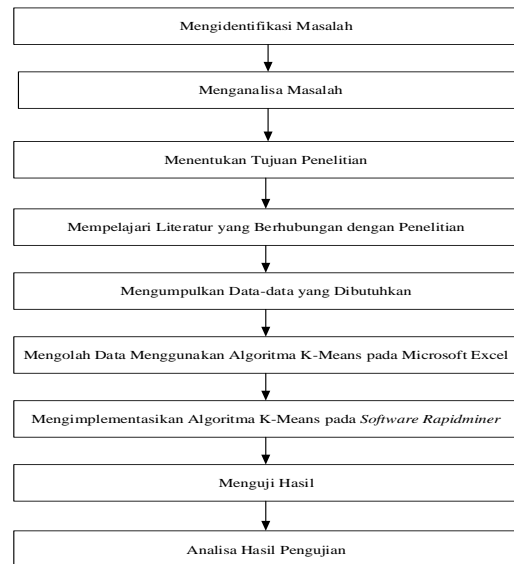
Gambar 1. KDD Proses

Data mining memiliki dua jenis model yaitu model prediktif yang digunakan untuk memprediksi suatu nilai dimasa depan dan model deskriptif yang digunakan untuk menemukan pola yang menggambarkan data sehingga bisa dimengerti manusia [15]. Teknik data mning memiliki beberapa teknik seperti klasifikasi, klasterisasi, prediksi dan asosiasi [16]. K-means Clustering merupakan salah satu algoritma klasterisasi yang digunakan dalam data mining [17].

Algoritma K-means digunakan dalam clustering, proses utamanya adalah menemukan mean dari elemen-elemen setiap subset cluster dan menetapkan sebagai pusat cluster. Cara kerja utama K-means adalah membagi elemen-elemen menjadi cluster setelah perulangan iterasi, jika tidak ada perpindahan cluster elemen dari hasil iterasi sebelum maka iterasi akan berhenti, tetapi jika masih ada perubahan cluster elemen maka iterasi akan dilanjutkan [18]. Rapidminer merupakan salah satu software data mining [19]. Rapidminer juga bisa digunakan untuk algoritma data mining lainnya [20].

2. Metodologi Penelitian

Penelitian ini menggunakan langkah-langkah penelitian dapat dilihat pada Gambar 2.



Gambar 2. Kerangka Penelitian

2.1. Mengidentifikasi Masalah

Dalam mengidentifikasi masalah ini, terlebih dahulu kita harus dapat menentukan ruang lingkup dari masalahnya. Ruang lingkup ini ditentukan terlebih dahulu, guna untuk memperoleh solusi yang terbaik dari masalah yang diteliti. Dalam kasus ini, peneliti mendapatkan permasalahan mengenai pemilihan saham pada masa pandemi Covid-19. Seperti yang telah dijabarkan dilatar belakang bahwa saham memiliki resiko yang sangat tinggi.

2.2. Menganalisa Permasalahan

Setelah menentukan ruang lingkup permasalahan, selanjutnya peneliti akan melakukan proses analisa masalah. Proses analisa ini berfungsi untuk dapat lebih memahami masalah yang diteliti sesuai dengan ruang lingkup yang sudah ditentukan. Nantinya, hasil pengujian ini diharapkan mampu memberikan pilihan saham pada masa pandemi Covid-19 untuk proses pengambilan keputusan untuk melakukan investasi.

2.3. Menentukan Tujuan Dari Penelitian

Pada langkah ini, ditentukan tujuan dari penelitian yang ingin dicapai. Yang mana, dengan penelitian ini diharapkan mampu mengurangi resiko dalam berinvestasi saham pada masa pandemi Covid-19

2.4. Mempelajari Literatur yang digunakan

Untuk mewujudkan tujuan penelitian, maka peneliti mempelajari beberapa literatur-literatur yang digunakan sesuai dengan penelitian yang diteliti. Literatur yang digunakan dalam penelitian ini berupa artikel ilmiah data mining tentang Algoritma k-means, serta artikel pendukung lainnya.

2.5. Mengumpulkan data-data yang dibutuhkan

Adapun beberapa metode dalam pengumpulan data yang dilakukan peneliti adalah sebagai berikut:

a. Dokumen

Melakukan pengumpulan data bersifat documenter dari internet. Peneliti langsung mengunduh dari website PT. Bursa Efek Indonesia dan yahoo finance.

b. Studi Pustaka

Mengumpulkan dan mempelajari artikel ilmiah yang berkaitan tentang penelitian yang diangkat.

2.6. Mengolah Data menggunakan Algoritma K-Means pada Microsoft Excel

Pada tahap ini, data yang diperoleh di tahap pengumpulan data diproses dan diolah dengan menggunakan algoritma K-Means menggunakan Microsoft excel.

2.7. Mengimplementasi Algoritma K-Means pada Software Rapidminer

Pada tahap ini dilakukan implementasi terhadap data pada software Rapidminer. Implementasi ini bertujuan untuk melakukan validasi perhitungan dengan microsoft excel.

2.8. Menguji Hasil

Langkah selanjutnya melakukan pengujian hasil dengan menggunakan algoritma k-means guna dapat mengetahui tingkat akurasi yang dihasilkan. Adapun beberapa tahapan-tahapan yang dilakukan dalam melakukan pengujian sistem adalah:

a. Menguji hasil dengan menggunakan algoritma K-Means pada Microsoft excel.

b. Menguji hasil dengan menggunakan algoritma K-Means pada software Rapidminer.

2.9. Melakukan Analisa Hasil Pengujian

Melakukan analisa hasil yang telah diperoleh dari data yang telah diolah dengan menggunakan algoritma K-Means. Sehingga peneliti dapat melakukan penarikan kesimpulan dari hasil analisa penelitian.

3. Hasil dan Pembahasan

3.1. Tahapan Analisa dan Perancangan

Dalam melakukan analisa dan implementasi pada software Rapidminer perlu dilakukan sebuah analisa secara terstruktur.

Data yang digunakan adalah data histori saham-saham yang termasuk dalam saham LQ-45 dari Januari 2019 sampai dengan Desember 2021. Data histori saham LQ-45 diunduh dari situs bursa efek Indonesia,

sedangkan data histori saham diunduh dari finance.yahoo.com. Data histori saham seperti open, high, low, close. Data histori tersebut akan diolah untuk menentukan perubahan harga per periode waktu sehingga bisa dilakukan klusterisasi.

3.2. Data Selection

Data selection adalah pemilihan yang dilakukan terhadap banyaknya data yang sesuai dengan kebutuhan penelitian. Pemilihan data ini dilakukan untuk mendapatkan data uji. Saham LQ-45 selalu mengalami perubahan perperiode sesuai dengan hasil evaluasi oleh PT. Bursa Efek Indonesia, dikarenakan ada perubahan daftar saham LQ-45 dari Januari 2019 sampai dengan Desember 2021, penelitian ini memilih saham-saham yang selalu ada didalam daftar saham LQ-45 dari Januari 2019 sampai dengan Desember 2021. Saham-saham tersebut berjumlah sebanyak 34 saham dengan frekuensi transaksi bulanan yang disajikan pada Tabel 1.

Tabel.1. Data Daftar Saham LQ-45

No.	Kode	Nama Saham
1	ADRO	Adaro Energy Tbk.
2	AKRA	AKR Corporindo Tbk.
3	ANTM	Aneka Tambang Tbk.
4	ASII	Astra International Tbk.
5	BBCA	Bank Central Asia Tbk.
6	BBNI	Bank Negara Indonesia (Persero) Tbk.
7	BBRI	Bank Rakyat Indonesia (Persero) Tbk.
8	BBTN	Bank Tabungan Negara (Persero) Tbk.
9	BMRI	Bank Mandiri (Persero) Tbk.
10	BSDE	Bumi Serpong Damai Tbk.
11	CPIN	Charoen Pokphand Indonesia Tbk
12	ERAA	Erajaya Swasembada Tbk.
13	EXCL	XL Axiata Tbk.
14	GGRM	Gudang Garam Tbk.
15	HMSP	H.M. Sampoerna Tbk.
16	ICBP	Indofood CBP Sukses Makmur Tbk.
17	INCO	Vale Indonesia Tbk.
18	INDF	Indofood Sukses Makmur Tbk.
19	INKP	Indah Kiat Pulp & Paper Tbk.
20	INTP	Indocement Tungal Prakarsa Tbk.
21	ITMG	Indo Tambangraya Megah Tbk.
22	JSMR	Jasa Marga (Persero) Tbk.
23	KLBF	Kalbe Farma Tbk.
24	MNCN	Media Nusantara Citra Tbk.
25	PGAS	Perusahaan Gas Negara (Persero) Tbk.
26	PTBA	Bukit Asam Tbk.
27	PTPP	PP (Persero) Tbk.
28	PWON	Pakuwon Jati Tbk.
29	SMGR	Semen Indonesia (Persero) Tbk.
30	TKIM	Pabrik Kertas Tjiwi Kimia Tbk.
31	TLKM	Telekomunikasi Indonesia (Persero) Tbk.
32	UNTR	United Tractors Tbk.
33	UNVR	Unilever Indonesia Tbk.
34	WIKA	Wijaya Karya (Persero) Tbk.

3.3. Pembersihan Data

Data yang telah dipilih pada proses data selection kemudian dilakukan proses pembersihan data. Proses pembersihan data bertujuan untuk melakukan validasi

data, seperti adanya data yang salah atau data yang tidak dibutuhkan dilakukan terlebih dahulu pembersihan data dengan cara pembenaran ataupun pembuangan data. Data yang dibutuhkan adalah data perubahan antara harga close dengan harga open per satuan waktu. Perubahan harga dihitung dengan Persamaan (1).

$$C = \frac{HC - HO}{HO} HO \quad (1)$$

Dimana C adalah perubahan harga, C dihitung dengan cara HC (harga close) dikurangi HO (harga open) kemudian dibagi dengan HO (harga open). Berdasarkan dari persamaan diatas akan didapatkan nilai perubahan harga saham per periode.

3.4. Transformasi Data

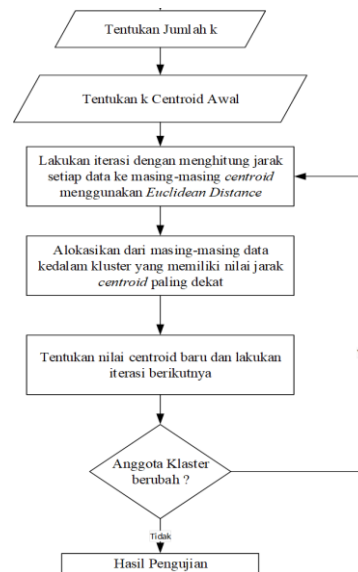
Transformasi data dilakukan bertujuan untuk mengubah data agar dapat diolah dengan menggunakan metode K-Means. Proses transformasi yang dilakukan adalah dengan menggunakan perubahan harga per periode waktu pada variabel data. 3.5 Melakukan Perhitungan dengan K-Means Clustering

Algoritma K-Means adalah pengelompokan data yang berdasarkan jarak data ke titik pusat dari kluster (centroid) yang terdekat dengan data yang merupakan algoritma klusterisasi. Adapun tujuan dari algoritma ini yaitu mengelompokkan data dengan memaksimalkan kemiripan data atau pola dalam satu kluster dan meminimalkan kemiripan data antar kluster. Fungsi jarak merupakan ukuran kemiripan yang digunakan dalam kluster, sehingga hasil yang maksimal dari kemiripan data berdasarkan jarak terpendek antara data terhadap titik centroid. Langkah-langkah dalam melakukan Clustering dengan metode K-Means adalah sebagai berikut:

Tahapan proses analisa Metode Algoritma K-Means disajikan pada Gambar 2 dengan urutan sebagai berikut:

- Menentukan jumlah k.
- Menentukan k centroid awal secara acak.
- Lakukan iterasi dengan menghitung jarak setiap data ke masing-masing centroid menggunakan Euclidean Distance.
- Alokasikan dari masing-masing data kedalam kluster yang memiliki nilai jarak centroid paling dekat.
- Tentukan nilai centroid baru dan lakukan iterasi berikutnya.

- Jika posisi data setiap kluster mengalami perubahan, maka ulangi langkah tiga.
- Jika posisi data setiap kluster tidak berubah maka hentikan iterasi dan sudah dapat hasil pengujian.



Gambar 3. Proses K-Means Clustering

3.5.1 Menentukan Jumlah Kluster

Untuk menggunakan algoritma K-Means perlu ditentukan jumlah k atau yang digunakan untuk melakukan klasifikasi data baru. Banyaknya jumlah k yang dipilih adalah 3 (K01, K02, K03). K01 merupakan kluster uptrend, K02 merupakan kluster no trend, K03 merupakan kluster downtrend. Penentuan nilai k dipertimbangkan berdasarkan banyaknya data yang ada yang dibentuk oleh data. Semakin banyak data yang ada, angka k yang dipilih sebaiknya semakin banyak. Namun, semakin kecil data, angka k yang dipilih sebaiknya semakin sedikit.

3.5.2. Menentukan Centroid Setiap Kluster

Iterasi pertama dilakukan dengan menggunakan centroid secara acak untuk masing-masing cluster. Untuk menentukan k centroid diambil data secara acak. Data TK3 merupakan centroid untuk kluster K01, data TK31 merupakan centroid untuk kluster K02, TK15 merupakan centroid untuk kluster K03.

3.5.3 Menghitung Jarak Setiap Data Ke Masing-Masing Centroid

Perhitungan jarak setiap data ke centroid dilakukan dengan menggunakan persamaan Euclidean Distance. Data yang digunakan adalah data hasil transformasi pada Tabel 2.

3.5.4. Mengelompokkan Data ke Kluster

Data yang memiliki jarak terkecil diantara 3 centroid, maka akan dikelompokkan kedalam kluster centroid tersebut. Perhitungan mulai dilakukan dari iterasi-1 sampai dengan iterasi-n. Iterasi akan berhenti jika tidak ada perubahan kluster data iterasi terakhir dengan iterasi sebelumnya. Setiap melakukan iterasi maka akan memiliki nilai centroid yang baru.

3.5.5 Hasil Pengujian

Selanjutnya dari pengelompokkan iterasi 4 diatas maka cluster baru sama dengan cluster sebelumnya, maka iterasi dihentikan. Berdasarkan hasil iterasi 1, 2, 3 dan 4 diatas, maka dapat hasil perhitungan pada Tabel 2.

Tabel 2. Perhitungan K-Means dari Manual

Kode	Ticker	Manual		Sistem	
		Cluster	Trend	Cluster	Trend
TK1	ADRO	2	No Trend	cluster_2	Down Trend
TK2	AKRA	2	No Trend	cluster_2	Down Trend
TK3	ANTM	1	Up Trend	cluster_1	No Trend
TK4	ASII	3	Down trend	cluster_2	Down Trend
TK5	BBCA	2	No Trend	cluster_2	Down Trend
TK6	BBNI	2	No Trend	cluster_0	Up Trend
TK7	BBRI	2	No Trend	cluster_2	Down Trend
TK8	BBTN	2	No Trend	cluster_0	Up Trend
TK9	BMRI	2	No Trend	cluster_2	Down Trend
TK10	BSDE	2	No Trend	cluster_0	Up Trend
TK11	CPIN	3	Down trend	cluster_2	Down Trend
TK12	ERAA	1	Up Trend	cluster_1	No Trend
TK13	EXCL	2	No Trend	cluster_2	Down Trend
TK14	GGRM	3	Down trend	cluster_2	Down Trend
TK15	HMSF	3	Down trend	cluster_2	Down Trend
TK16	ICBP	2	No Trend	cluster_2	Down Trend
TK17	INCO	1	Up Trend	cluster_1	No Trend
TK18	INDF	2	No Trend	cluster_2	Down Trend
TK19	INKP	1	Up Trend	cluster_1	No Trend
TK20	INTP	3	Down trend	cluster_0	Up Trend
TK21	ITMG	2	No Trend	cluster_2	Down Trend
TK22	JSMR	2	No Trend	cluster_0	Up Trend
TK23	KLBF	2	No Trend	cluster_2	Down Trend
TK24	MNCN	2	No Trend	cluster_2	Down Trend
TK25	PGAS	2	No Trend	cluster_0	Up Trend
TK26	PTBA	2	No Trend	cluster_2	Down Trend
TK27	PTPP	2	No Trend	cluster_0	Up Trend
TK28	PWON	2	No Trend	cluster_0	Up Trend
TK29	SMGR	3	Down trend	cluster_2	Down Trend
TK30	TKIM	1	Up Trend	cluster_1	No Trend
TK31	TLKM	2	No Trend	cluster_2	Down Trend
TK32	UNTR	2	No Trend	cluster_2	Down Trend
TK33	UNVR	2	No Trend	cluster_2	Down Trend
TK34	WKA	2	No Trend	cluster_0	Up Trend

Kedua hasil perhitungan dibandingkan dalam mengidentifikasi keakuratan pengolahan. Tingkat akurasi ini menjadi landasan dalam analisa trend saham dalam perubahan harga berdasarkan Persamaan (1). Yang disajikan pada Tabel 3.

Tabel 3. Rata-rata Perubahan Harga

Ticker	Perubahan (%)
ADRO	2
AKRA	0
ANTM	4
ASII	-1
BBCA	1
BBNI	0
BBRI	0
BBTN	1
BMRI	0
BSDE	0
CPIN	0
ERAA	2
EXCL	2
GGRM	-2
HMSF	-4
ICBP	-1
INCO	1
INDF	0
INKP	0
INTP	-1
ITMG	1
JSMR	1
KLBF	0
MNCN	2
PGAS	0
PTBA	-1
PTPP	0
PWON	0
SMGR	-1
TKIM	1
TLKM	0
UNTR	0
UNVR	-2
WKA	0

Tabel 3 menyajikan perubahan hasil perhitungan dalam mengidentifikasi keakuratan pengolahan. Tingkat akurasi ini menjadi landasan dalam analisa trend saham dalam perubahan harga.

4. Kesimpulan

Penggunaan Algoritma K-Means Clustering memiliki validasi sebesar 32% terhadap persentase rata-rata perubahan harga. Hasil ini menjadi koreksi terhadap perubahan harga yang terjadi dilapangan untuk periode berikutnya. Sehingga penelitian ini dapat menjadi rujukan dalam analisis harga saham untuk periode yang akan datang.

Daftar Rujukan

- [1] Mukhsin, M. (2020). Peranan Teknologi Informasi dan Komunikasi Menerapkan Sistem Informasi Desa dalam Publikasi Informasi Desa di Era Globalisasi. Teknokon, 3(1), 7-15. <https://doi.org/10.31943/teknokon.v3i1.43>
- [2] Mubarak, F. S. Pemanfaatan New Media untuk Efektivitas Komunikasi di Era Pandemi. <http://dx.doi.org/10.30659/jikm.v10i1.20302>
- [3] Fitriani, Y. (2021). Analisa Pemanfaatan Aplikasi Keuangan Online Sebagai Media Untuk Mengelola Atau Memanajemen Keuangan. Journal of Information System, Applied, Management, Accounting and Research, 5(2), 454-461. <https://doi.org/10.52362/jisamar.v5i2.432>

- [4] Mangarey, R. B. (2021). Pengaruh Kebijakan Dividen dan Profitabilitas Terhadap Nilai Perusahaan Pada Perusahaan LQ 45 Yang Terdaftar Di Bursa Efek Indonesia. *Jurnal EMBA: Jurnal Riset Ekonomi, Manajemen, Bisnis dan Akuntansi*, 9(1). <https://doi.org/10.35794/emba.v9i1.32185>
- [5] Balqis, B. (2021). Determinasi Earning Per Share dan Return Saham: Analisis Return On Asset, Debt To Equity Ratio, dan Current Ratio. *Jurnal Ilmu Manajemen Terapan*, 2(5), 665-675. <https://doi.org/10.31933/jimt.v2i5.511>
- [6] Alhazami, L. (2020). The Valuasi Saham yang Masih Layak untuk Dikoleksi di Bursa Efek Indonesia (BEI) pada Saat Pandemi Covid-19. *Jurnal Ilmiah Akuntansi Dan Keuangan*, 9(2), 139-149. <https://doi.org/10.32639/jiak.v9i2.526>
- [7] Sumarni, Y. (2020). Pandemi Covid-19: Tantangan ekonomi dan bisnis. *Al-Intaj: Jurnal Ekonomi Dan Perbankan Syariah*, 6(2), 46-58. <http://dx.doi.org/10.29300/aij.v6i2.3358>
- [8] Purnama, J., & Juliana, A. (2020). Analisa Prediksi Indeks Harga Saham Gabungan Menggunakan Metode Arima. *Cakrawala Management Business Journal*, 2(2), 454-468. <https://doi.org/10.30862/cm-bj.v2i2.51>
- [9] Zullah, V. S., Prastuti, M., & Santoso, B. H. (2022). Peramalan Harga Saham Perbankan Menggunakan Model VAR dan Analisis Teknikal. *Jurnal Sains dan Seni ITS*, 11(1), D110-D115. <http://dx.doi.org/10.12962/j23373520.v11i1.63225>
- [10] Manullang, S. D., Buulolo, E., & Lubis, I. (2020). Implementasi Data Mining Dalam Memprediksi Jumlah Pinjaman Dengan Algoritma C4. 5 Pada Kopdit CU Damai Sejahtera. *Jurnal Sistem Komputer dan Informatika (JSON)*, 1(3), 265-272. <http://dx.doi.org/10.30865/json.v1i3.2153>
- [11] Khotimah, K. (2020). PENGELOLAAN BIG DATA PERPUSTAKAAN DENGAN TEKNIK DATA MINING. *Sienna*, 1(2), 64-70. <https://doi.org/10.47637/sienna.v1i2.349>
- [12] Rejeki, W. S. (2021). Implementasi Algoritma C. 45 Pada Data Pengolahan Limbah Kelapa Sawit. *BEES: Bulletin of Electrical and Electronics Engineering*, 2(2), 40-46. <https://doi.org/10.47065/bees.v2i2.696>
- [13] Nuraliza, H., Pratiwi, O. N., & Hamami, F. (2022). Analisis Sentimen IMbD Film Review Dataset Menggunakan Support Vector Machine (SVM) dan Seleksi Feature Importance. *Jurnal Mirai Management*, 7(1), 1-17. <https://doi.org/10.37531/mirai.v7i1.2222>
- [14] Rahmi, A. N., & Mikola, Y. A. (2021). Implementasi Algoritma Apriori Untuk Menentukan Pola Pembelian Pada Customer (Studi Kasus: Toko Bakoel Sembako). *Information System Journal*, 4(1), 14-19. <https://doi.org/10.24076/infosjournal.2021v4i1.561>
- [15] Wu, W. T., Li, Y. J., Feng, A. Z., Li, L., Huang, T., Xu, A. D., dan Lyu, J. (2021). Data Mining in clinical big data: the frequently used databases, steps, and methodological models. *Military Medical Research*, 8(1), 1-12. <https://doi.org/10.1186/s40779-021-00338-z>
- [16] Pasaribu, A. F. O. (2021). Analisis Pola Menggunakan Metode C4. 5 Untuk Peminatan Jurusan Siswa Berdasarkan Kurikulum (Studi Kasus: Sman 1 Natar). *Jurnal Teknologi Dan Sistem Informasi*, 2(1), 80-85. <https://doi.org/10.33365/jtsi.v2i1.393>
- [17] Suputra, W. A. (2021). Klasterisasi Hasil Ujian Nasional SMA/MA dengan Algoritma K-Means. *Wahana Matematika dan Sains: Jurnal Matematika, Sains, dan Pembelajarannya*, 15(1), 22-30. <https://doi.org/10.23887/wms.v15i1.25380>
- [18] Lu, W. (2020). Improved K-means clustering algorithm for big Data Mining under Hadoop parallel framework. *Journal of Grid Computing*, 18(2), 239-250. <https://doi.org/10.1007/s10723-019-09503-0>
- [19] Sudarsono, B. G., Leo, M. I., Santoso, A., & Hendrawan, F. (2021). Analisis Data Mining Data Netflix Menggunakan Aplikasi Rapid Miner. *JBASE-Journal of Business and Audit Information Systems*, 4(1). <http://dx.doi.org/10.30813/jbase.v4i1.2729>
- [20] Ordila, R., Wahyuni, R., Irawan, Y., & Sari, M. Y. (2020). Penerapan Data Mining Untuk Pengelompokan Data Rekam Medis Pasien Berdasarkan Jenis Penyakit Dengan Algoritma Clustering (Studi Kasus: Poli Klinik Pt. Inecda). *Jurnal Ilmu Komputer*, 9(2), 148-153. <https://doi.org/10.33060/JIK/2020/Vol9.Iss2.181>