

PENERAPAN DATA MINING DALAM PENGELOMPOKAN SAHAM MELALUI RASIO KEUANGAN PADA SAHAM PAPAN UTAMA BURSA EFEK INDONESIA MENGGUNAKAN METODE K-MEANS CLUSTERING

Nopri Jumairi, Jasmir, Benni Purnama

Magister Sistem Informasi, Universitas Dinamika Bangsa Jambi

Jl. Jend. Sudirman, The Hok, Kota Jambi, Indonesia

Noprijumairi@gmail.com

ABSTRAK

Investasi saham di Bursa Efek Indonesia kian menarik perhatian, terutama dari kalangan generasi muda. Akan tetapi, rendahnya pemahaman tentang keuangan mengakibatkan banyak investor sulit untuk memilih saham yang tepat, sehingga berisiko menderita kerugian. Salah satu metode untuk mendukung investor dalam proses pengambilan keputusan adalah dengan mengelompokkan saham sesuai dengan rasio keuangan. Studi ini menggunakan metode K-Means Clustering untuk mengelompokkan saham di Papan Utama Bursa Efek Indonesia berdasarkan rasio keuangan, yaitu *Price Earning Ratio* (PER), *Earning Per Share* (EPS), *Book Value Per Share* (BVPS), dan *Price to Book Value* (PBV). Temuan riset menunjukkan bahwa pendekatan ini mengelompokkan saham ke dalam tiga kategori risiko: Centroid 1 (risiko rendah) memiliki 12 saham, Centroid 2 (risiko sedang) terdiri dari 223 saham, dan Centroid 3 (risiko tinggi) mencakup 3 saham. Perbandingan dengan RapidMiner menampilkan variasi jumlah saham di setiap kategori, dengan hasil pengelompokan yang berbeda. Studi ini menemukan bahwa teknik *K-Means Clustering* dapat memfasilitasi investor dalam memahami profil risiko saham secara lebih mendalam, yang pada gilirannya mendukung keputusan investasi yang lebih bijaksana dan mengurangi kemungkinan kerugian.

Kata Kunci : Saham, Data Mining, K-Means Clustering, Rasio Keuangan, RapidMiner

1. PENDAHULUAN

Investasi saham semakin populer di kalangan banyak orang, terutama generasi muda, sebagai salah satu instrumen finansial yang memberikan peluang keuntungan lebih besar daripada jenis investasi lainnya. Namun, dalam praktiknya, investor baru sering menemui berbagai kesulitan dalam membuat keputusan investasi yang terbaik, yang sebagian besar disebabkan oleh rendahnya tingkat literasi keuangan dan minimnya pemahaman mengenai metode analisis saham yang benar. Keterbatasan dalam pemahaman analisis fundamental, termasuk penggunaan rasio keuangan untuk mengevaluasi kinerja perusahaan, membuat banyak investor cenderung mengandalkan tren pasar atau saran dari pihak lain tanpa mempertimbangkan aspek fundamental perusahaan yang mereka investasikan. Ini dapat meningkatkan paparan terhadap risiko kerugian, khususnya bagi investor yang hanya mengikuti sentimen pasar secara spekulatif tanpa dukungan analisis yang kuat.

Kurangnya pemahaman yang cukup dalam memilih saham dengan tepat dapat menimbulkan efek negatif yang signifikan, baik bagi investor perorangan maupun untuk stabilitas pasar modal secara keseluruhan. Salah satu dampak utama dari kesalahan dalam memilih saham adalah besarnya kerugian yang terjadi akibat keputusan investasi yang tidak didukung oleh analisis keuangan yang mendetail. Mayoritas investor pemula biasanya membeli saham berdasarkan tren harga tanpa menyadari apakah valuasi saham itu sudah mencerminkan kondisi fundamental perusahaan atau hanya merupakan produk dari spekulasi pasar. Di samping itu,

kurangnya pemahaman tentang strategi diversifikasi portofolio juga bisa menambah risiko sistemik, karena para investor mungkin terfokus pada saham-saham yang memiliki volatilitas tinggi tanpa adanya langkah mitigasi risiko yang cukup. Dalam jangka panjang, jika sebagian besar investor mengalami kerugian signifikan karena kurangnya pemahaman ini, maka kepercayaan terhadap pasar modal dapat berkurang, menurunkan tingkat likuiditas, serta berdampak pada stabilitas harga saham secara keseluruhan. Oleh karena itu, diperlukan pendekatan yang berbasis data untuk membantu investor memahami karakteristik saham dengan objektif, sehingga mereka dapat membuat keputusan investasi yang lebih rasional dan berbasis data.

Sejalan dengan kemajuan teknologi di sektor keuangan dan analisis data, penggunaan metode data mining bisa menjadi alternatif yang efisien untuk menemukan pola-pola dalam data keuangan yang rumit. Salah satu cara yang bisa diterapkan untuk mengelompokkan saham berdasarkan sifat fundamentalnya adalah metode *K-Means Clustering*, yang merupakan teknik *unsupervised learning* dalam *machine learning* yang dapat mengklasifikasikan data ke dalam beberapa kategori berdasarkan kesamaan pola dalam *dataset* yang ada.

Pemilihan metode *K-Means Clustering* dalam penelitian ini didasari oleh beberapa kelebihan utama yang menjadikannya alat analisis yang sesuai untuk mengelompokkan saham di Bursa Efek Indonesia. Pertama, pendekatan ini memiliki kapasitas untuk mengklasifikasikan data dalam skala besar dengan efisien, sehingga sangat sesuai untuk digunakan

dalam analisis data keuangan yang kompleks dan multivariabel. Kedua, *K-Means Clustering* memberikan kebebasan dalam memilih jumlah kluster yang sesuai dengan tujuan analisis, sehingga dalam konteks penelitian ini, saham dapat dikelompokkan ke dalam kategori risiko rendah, sedang, dan tinggi berdasarkan rasio keuangan yang diterapkan. Ketiga, hasil pengelompokan yang diperoleh dari algoritma ini bisa dengan mudah dipahami oleh investor, sehingga dapat mendukung mereka dalam mengenal karakteristik saham yang beragam di Papan Utama Bursa Efek Indonesia. Di samping itu, teknik ini juga bisa digunakan dengan berbagai aplikasi perangkat lunak, seperti *RapidMiner*, *Python*, dan *Excel*, yang membuat proses analisis berlangsung lebih cepat, tepat, dan dapat dibandingkan dengan metode perhitungan secara manual.

Studi ini bertujuan menerapkan metode *K-Means Clustering* untuk mengelompokkan saham yang terdaftar di Papan Utama Bursa Efek Indonesia berdasarkan rasio keuangan utama, seperti *Price Earning Ratio* (PER), *Earning Per Share* (EPS), *Book Value Per Share* (BVPS), dan *Price to Book Value* (PBV). Selanjutnya, studi ini juga berfokus pada penilaian efektivitas *K-Means Clustering* dalam mengelompokkan saham berdasarkan risiko investasi dan membandingkan hasil perhitungan manual dengan analisis yang dilakukan menggunakan perangkat lunak *RapidMiner* untuk mengevaluasi tingkat akurasi metode ini dalam mengklasifikasikan saham yang dapat dijadikan sebagai referensi investasi bagi para investor. Dengan penerapan pengelompokan saham melalui metode ini, diharapkan para investor dapat memahami lebih baik karakteristik saham yang mereka pilih, sehingga dapat menurunkan risiko investasi dan meningkatkan peluang keuntungan melalui pendekatan investasi yang lebih sistematis, berbasis data, dan dapat dipertanggungjawabkan secara ilmiah.

2. TINJAUAN PUSTAKA

2.1. Data Mining

Data mining merupakan suatu proses yang ditujukan untuk mengambil nilai tambah berupa informasi yang sebelumnya tidak terungkap dari sebuah basis data. Proses ini melibatkan pengidentifikasian pola data untuk mengkonversi data menjadi informasi yang lebih berharga, dengan cara mengenali dan mengekstrak pola-pola yang signifikan atau menarik dalam database [1]

2.2. Clustering

Clustering adalah teknik data mining yang tidak terarah (*unsupervised*), yang berarti metode ini digunakan tanpa pelatihan (*training*), tanpa supervisi (mentor), dan tanpa memerlukan target hasil. [2]

Metode penelitian yang digunakan adalah analisis deskriptif dengan pendekatan kuantitatif. Ini menunjukkan bahwa penelitian berfokus pada analisis data numerik (angka), dengan tujuan untuk mendapatkan pemahaman yang jelas mengenai

kondisi yang didasarkan pada data yang telah dikumpulkan. Proses ini mencakup penyajian, pengumpulan, dan analisis data guna menghasilkan informasi baru yang dapat digunakan dalam menganalisis masalah yang sedang diteliti [3]

Tahap berikutnya adalah menghitung jarak dari setiap titik yang akan dikelompokkan ke masing-masing *centroid* yang ada, dan mengelompokkan titik tersebut berdasarkan jarak terdekat ke *centroid*. Setelah semua titik dikelompokkan, tahap ini dianggap selesai. Selanjutnya, *centroid* baru dihitung sebagai pusat dari setiap kelompok yang terbentuk. Setelah memperoleh *centroid* yang baru, pengelompokan diuji kembali berdasarkan *centroid* tersebut. Proses ini diulang hingga *centroid* stabil dan tidak mengalami perubahan signifikan. Penentuan kelompok *cluster* dilakukan dengan menghitung jarak setiap data dari *centroid* awal, seperti data ke-1 dan data ke-2, menggunakan jarak Euclidean[4].

Proses ini dilakukan dengan menerapkan rumus berikut:

$$D(x, y) = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2} \dots \dots \dots (1)$$

Langkah-langkah algoritma untuk melaksanakan *K-Means clustering* adalah sebagai berikut:

- Pilih K buah titik *centroid* secara acak
- Kelompokkan data sehingga terbentuk K buah *cluster* dengan titik *centroid* dari setiap *cluster* merupakan titik yang telah dipilih sebelumnya
- Perbaharui nilai titik *centroid*
- Ulangi langkah 2 dan 3 hingga nilai *centroid* tidak lagi berubah.

$$\text{Di mana: } \mu_k = \frac{1}{N_k} \sum_{q=1}^{N_k} X_q \dots \dots \dots (2)$$

μ_k = titik *centroid* dari *cluster* ke-K

N_k = banyaknya data pada *cluster* ke-K

x_q = data ke-q pada *cluster* ke-K

Untuk menghitung jarak ke suatu *centroid* masing-masing *Centroid*, misalkan data (x,y), *centroid* M1: (a1,b1), *centroid* M2: (a2,b2), dan *centroid* M3: (a3,b3)

$$dM1 = \sqrt{(x - a1)^2 + (y - b1)^2} = ? \dots \dots \dots (3)$$

$$dM2 = \sqrt{(x - a2)^2 + (y - b2)^2} = ? \dots \dots \dots (4)$$

$$dM3 = \sqrt{(x - a3)^2 + (y - b3)^2} = ? \dots \dots \dots (5)$$

2.3. Pasar Modal

Pasar modal berperan sebagai penghubung penting bagi investor dan menjadi fondasi perekonomian. Pasar ini menghubungkan individu atau entitas yang membutuhkan dana dengan individu atau entitas yang memiliki surplus dana, dan juga memfasilitasi pengalokasian dana yang efisien. [5]

Dengan pasar modal, investor memiliki pilihan beragam opsi investasi alternatif yang menawarkan hasil terbaik. Untuk investor yang mengevaluasi perusahaan, laporan keuangan yang diterbitkan memberikan pemahaman mengenai pertumbuhan kebutuhan perusahaan dan prestasi yang telah diperoleh [6]

2.4. Saham

Saham merupakan bukti kepemilikan dalam suatu perusahaan yang mencantumkan nilai nominal, nama perusahaan, serta secara jelas menetapkan hak dan tanggung jawab bagi pemilikinya. Saham adalah salah satu bentuk investasi modal atau hak kepemilikan atas dana di sebuah perusahaan [7]

Indeks Papan Utama dan Indeks Papan Pengembangan merupakan indeks harga saham yang menggambarkan pergerakan saham di setiap papan tersebut. Indeks ini diperoleh dengan menghitung semua saham yang terdaftar di setiap papan, kecuali saham-saham tertentu yang tidak termasuk dalam perhitungan berdasarkan pertimbangan khusus. [8]

2.5. Rasio Keuangan

Investor memerlukan data historis pergerakan saham di pasar, baik secara individu, kelompok, maupun kombinasi, untuk menentukan saham yang akan dijadikan objek investasi. Target utama setiap investor adalah mendapatkan laba, salah satunya melalui capital gain, yaitu perbedaan antara harga beli dan harga jual saham, di mana harga jual lebih tinggi dibandingkan harga beli [9]

- a. *Price Earning Ratio* (PER) berguna untuk mengevaluasi bagaimana pasar menilai kinerja saham perusahaan menurut laba per saham. Rasio yang lebih tinggi menunjukkan bahwa kinerja perusahaan semakin baik, sedangkan *Price Earning Ratio* (PER) yang terlalu tinggi dapat menunjukkan bahwa harga saham sudah terlalu mahal atau tidak masuk akal [10]

- b. *Earning per share* (EPS) merupakan bentuk distribusi keuntungan yang diberikan kepada pemegang saham dari setiap lembar saham yang dimiliki *Earning per share* (EPS) adalah salah satu rasio keuangan yang sering dipakai oleh investor untuk menganalisa kemampuan perusahaan dalam menghasilkan laba [11]
- c. *Book Value Per Share* (BVPS) merupakan rasio yang dipakai untuk membandingkan ekuitas pemegang saham dengan total saham yang beredar, diharapkan terus naik seiring dengan perbaikan kinerja perusahaan [11]
- d. *Price to Book Value Ratio* (PBV) Rasio ini membandingkan harga pasar saham dengan nilai buku per lembar saham, yang digunakan untuk menilai apakah harga saham dinilai terlalu tinggi atau terlalu rendah. [12]

2.6. RapidMiner

Rapidminer merupakan platform perangkat lunak yang fleksibel dan mudah diintegrasikan dengan berbagai bahasa pemrograman. Dikembangkan menggunakan Java, platform ini kompatibel dengan beragam sistem operasi dan menawarkan antarmuka pengguna grafis yang intuitif untuk membangun alur analisis data. Melalui antarmuka ini, pengguna dapat merancang proses analisis yang kemudian dikonversi menjadi file XML. File XML ini berfungsi sebagai blueprint yang mendefinisikan langkah-langkah analisis dan dieksekusi oleh RapidMiner untuk menghasilkan *output* yang diinginkan [13]

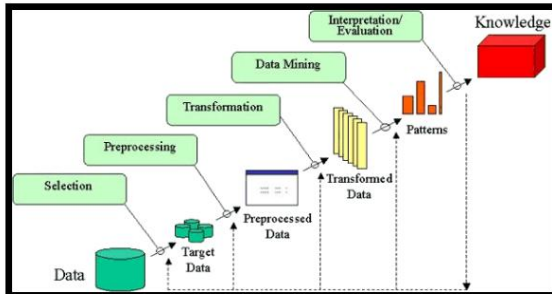
Tabel 1. Saham Papan Utama

No	Kode	Nama Perusahaan	EPS	PER	BVPS	PBV
1	AALI	Astra Agro Lestari Tbk.	565,37	11,85	11724,50	0,57
2	ABMM	ABM Investama Tbk.	1769,93	2,58	4255,86	1,07
3	ACES	Aspirasi Hidup Indonesia Tbk.	44,54	20,54	360,72	2,54
4	ADHI	Adhi Karya (Persero) Tbk.	34,48	8,06	1096,48	0,25
5	ADRO	Adaro Energy Indonesia Tbk.	931,05	4,20	3718,76	1,05
6	AGRO	Bank Raya Indonesia Tbk.	0,99	269,56	139,64	1,92
7	AISA	FKS Food Sejahtera Tbk.	2,02	65,39	103,98	1,27
8	AKRA	AKR Corporindo Tbk.	153,36	10,11	699,58	2,22
9	AMFG	Asahimas Flat Glass Tbk.	134,40	35,42	9771,28	0,49
10	AMRT	Sumber Alfaria Trijaya Tbk.	83,90	38,14	378,22	8,46
...	...	...	...	...	...	...
125	PNIN	Paninvest Tbk.	870,67	1,14	15661,92	0,06
126	PNLF	Panin Financial Tbk.	112,65	3,91	1935,57	0,23
127	PTBA	Bukit Asam Tbk.	546,19	5,75	1871,70	1,68
128	PTPP	PP (Persero) Tbk.	20,50	21,76	2442,52	0,18
129	PTRO	Petrosea Tbk.	190,39	70,64	3606,85	3,73
130	PUDP	Pudjiadi Prestige Tbk.	0,56	434,58	783,81	0,31
131	PWON	Pakuwon Jati Tbk.	49,46	10,41	473,33	1,09
...	...	...	...	...	...	...
233	BREN	Barito Renewables Energy Tbk.	16,76	428,00	75,05	95,60
234	MSTI	Mastersystem Infotama Tbk.	142,64	11,22	619,77	2,58
235	MSJA	Multi Spunindo Jaya Tbk.	19,78	17,09	178,21	1,90
236	ALII	Ancara Logistics Indonesia Tbk	9,59	45,24	40,11	10,82
237	GOLF	Intra Golfink Resorts Tbk.	3,09	71,89	389,07	0,57
238	BLES	Superior Prima Sukses Tbk.	16,98	14,13	85,01	2,82

Source : www.idx.co.id

Di tabel 1 terlihat bahwa terdapat 238 saham perusahaan yang tergolong dalam Papan Utama Bursa Efek Indonesia yang bersumber dari situs www.idx.co.id. Dengan rasio keuangan yang didasarkan pada laporan keuangan tahun 2023

3. METODE PENELITIAN



Gambar 1. Metode Knowledge Discovery using Database

Pada bagian ini memuat metode yang digunakan pada penelitian ini. Pada gambar 1, dapat dilihat tahapan proses menggunakan metode *Knowledge Discovery using Database*, dengan langkah sebagai berikut :

- a. Data Selection
Pemilihan Data bertujuan untuk menciptakan kumpulan data sasaran, memilih subset data, atau menitikberatkan pada variabel atau sampel data spesifik yang akan digunakan dalam analisis.
- b. Data Preprocessing
Dalam tahap *preprocessing* data, dilakukan pembersihan dari nilai yang hilang dan *outlier*. Selain itu, proses *cleaning* mencakup penghapusan duplikasi data, pemeriksaan inkonsistensi, dan perbaikan kesalahan seperti kesalahan tipografi.
- c. Data Transformation
Pada fase ini, ditetapkan variabel target dan faktor yang relevan. Mencari fitur yang tepat untuk menggambarkan data tergantung pada tujuan yang ingin dicapai. Proses ini meliputi perubahan data yang telah dipilih agar sesuai untuk analisis data *mining*.
- d. Data Mining
Proses penambangan data merupakan komponen penting dalam keseluruhan siklus Penemuan Pengetahuan dalam Basis Data (KDD). Pada fase penemuan pengetahuan, metode yang bisa digunakan mencakup klasifikasi, regresi, pengelompokan, dan lain-lain. Pemilihan algoritma data mining untuk mencari informasi harus disesuaikan dengan tujuan analisis dan karakteristik data yang ada, sesuai dengan proses pencarian pola atau informasi menarik dalam data yang telah dipilih, dengan menggunakan teknik atau metode tertentu.
- e. Data Interpretation
Penerjemahan pola-pola yang diperoleh dari data mining bertujuan untuk menyajikan informasi

dalam bentuk yang mudah dipahami oleh pihak-pihak yang berkepentingan. Tahap ini merupakan bagian dari proses KDD yang meliputi penilaian apakah pola atau informasi yang diperoleh sesuai dengan fakta atau hipotesis yang telah ada sebelumnya.

4. HASIL DAN PEMBAHASAN

4.1. Hasil Penelitian

4.1.1. Gambaran Umum BEI

Bursa Efek Indonesia (BEI) adalah salah satu institusi yang memiliki peran penting dalam ekonomi nasional dengan menawarkan infrastruktur untuk perdagangan efek. Perusahaan yang terdaftar di BEI dibagi menjadi beberapa papan pencatatan, yakni Papan Utama, Papan Pengembangan, dan Papan Akselerasi, yang masing-masing memiliki syarat berbeda yang harus dipenuhi oleh setiap emiten.

Papan Utama menjadi sorotan utama dalam studi ini karena terdiri dari perusahaan dengan nilai kapitalisasi pasar yang tinggi serta memiliki dasar keuangan yang lebih kokoh dibandingkan dengan perusahaan yang terdaftar di papan lainnya. Pergerakan harga saham di Papan Utama sangat dipengaruhi oleh kondisi ekonomi makro, laporan keuangan perusahaan, serta sentimen pasar yang sedang berlangsung. Sebagai akibatnya, penggunaan teknik data mining untuk mengelompokkan saham berdasarkan indikator keuangan menjadi sangat krusial dalam membantu investor untuk mengambil keputusan investasi yang lebih objektif dan terukur.

4.1.2. Pengumpulan Dan Seleksi Data

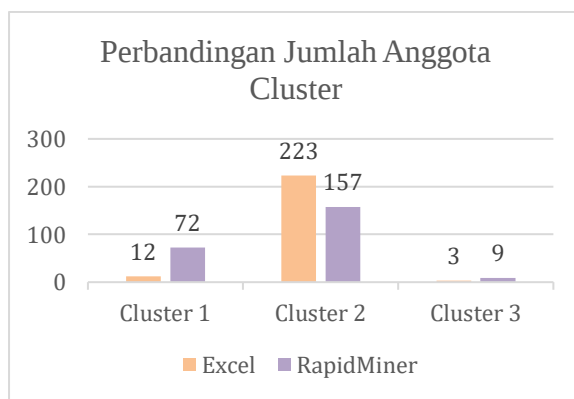
Dalam penelitian ini, data yang dipakai berasal dari laporan keuangan perusahaan yang telah diaudit dan diumumkan oleh Bursa Efek Indonesia (BEI). Dari keseluruhan perusahaan yang terdaftar di Papan Utama, dilakukan pemilihan berdasarkan ketersediaan informasi untuk empat rasio keuangan utama yang diterapkan dalam analisis kluster. Rasio-rasio keuangan yang dipilih mencakup *Earnings Per Share* (EPS) yang menilai laba bersih per lembar saham dan mencerminkan kapabilitas perusahaan dalam menghasilkan keuntungan bagi para pemegang saham. *Price to Book Value* (PBV) yang menunjukkan perbandingan antara harga pasar saham dan nilai buku perusahaan, sehingga dapat memberikan wawasan tentang valuasi suatu perusahaan di pasar saham. *Price Earning Ratio* (PER) yang mengukur rasio antara harga saham dan laba bersih per saham untuk menilai ekspektasi pasar terhadap pertumbuhan laba perusahaan. serta *Book Value Per Share* (BVPS) yang menggambarkan nilai aset bersih per lembar saham dan menjadi salah satu indikator dalam menilai kekuatan fundamental perusahaan. Setelah menjalani proses seleksi yang ketat, didapatkan 238 perusahaan yang memenuhi syarat kelengkapan data dan cocok untuk dianalisis lebih lanjut dalam proses *K-Means Clustering*.

4.1.3. Implementasi K-Means Clustering

Dalam studi ini, metode *K-Means Clustering* digunakan untuk mengelompokkan saham menurut kesamaan karakteristik keuangannya, yang dinilai melalui empat rasio utama, yaitu *Earnings Per Share* (EPS), *Price To Book Value* (PBV), *Price Earning Ratio* (PER), dan *Book Value Per Share* (BVPS). Proses pengelompokan dilaksanakan melalui beberapa langkah, dimulai dengan penentuan centroid awal, di mana tiga titik centroid awal diambil secara acak dari dataset yang ada. Selanjutnya, dihitung jarak Euclidean guna menentukan jarak antara setiap data saham dengan centroid awal yang telah ditentukan. Setelah perhitungan jarak, setiap saham dikelompokkan ke dalam kluster yang paling dekat dengan centroid, yang dikenal sebagai pengelompokan awal. Langkah selanjutnya adalah iterasi dan pembaruan centroid, di mana lokasi centroid diperbarui berdasarkan nilai rata-rata dari semua anggota dalam kluster yang terbentuk, dan proses ini dilakukan berulang kali hingga tercapai konvergensi, yaitu keadaan di mana tidak ada perubahan signifikan lagi dalam keanggotaan kluster.

4.1.4. Validasi Hasil dan Perbandingan

Untuk menilai ketepatan hasil dari perhitungan manual, metode yang serupa diterapkan dengan menggunakan perangkat lunak RapidMiner. Hasil analisis memperlihatkan variasi jumlah saham di setiap kluster sebagai berikut.



Gambar 2. Perbandingan Hasil

Pada gambar 2 diatas, terlihat perbandingan hasil antara perhitungan manual dan perhitungan yang menggunakan alat RapidMiner. Dalam perhitungan manual, terdapat tiga kelompok risiko investasi, di mana 12 saham berada di cluster 1 (risiko rendah), 223 saham di cluster 2 (risiko sedang), dan 3 saham di cluster 3 (risiko tinggi). Sedangkan jumlah cluster dalam perhitungan Rapidminer, 72 saham berada di cluster 1 (risiko rendah), 157 saham di cluster 2 (risiko sedang), dan 9 saham di cluster 3 (risiko tinggi).

Perbandingan antara perhitungan manual dan RapidMiner menunjukkan adanya perbedaan dalam jumlah saham yang masuk dalam masing-masing

kategori. RapidMiner menggunakan optimasi iteratif yang lebih kompleks dibandingkan perhitungan manual, yang memungkinkan pembagian kluster yang lebih seimbang. Dengan metode ini, investor dapat lebih mudah memilih saham berdasarkan profil risiko yang sesuai.

4.2. Pembahasan

4.2.1. Analisa Perbedaan Hasil Perhitungan

Hasil penelitian menunjukkan adanya variasi dalam jumlah saham yang dikelompokkan ke dalam tiap kluster, tergantung pada metode perhitungan yang diterapkan. Perbedaan ini muncul akibat beberapa faktor utama, salah satunya adalah inisialisasi centroid awal, di mana metode manual memungkinkan pemilihan centroid secara subjektif berdasarkan data yang ada, sedangkan RapidMiner menerapkan pendekatan optimasi otomatis untuk menentukan titik centroid awal yang lebih baik. Di samping itu, jumlah iterasi dalam metode manual adalah terbatas, sehingga hasil *clustering* dapat berbeda tergantung pada jumlah iterasi yang dilakukan, sedangkan RapidMiner dapat melakukan lebih banyak iterasi hingga mencapai stabilitas kluster secara optimal. Perbedaan lain berada pada penyebaran data di setiap kluster, di mana RapidMiner memiliki sistem otomatisasi yang bisa menyeimbangkan jumlah anggota di setiap kluster, sedangkan metode manual biasanya menghasilkan distribusi yang lebih tidak merata.

4.2.2. Hasil Clustering Untuk Keputusan Investasi

Dari hasil *clustering*, para investor bisa menggunakan informasi ini untuk merancang strategi investasi yang lebih efisien sesuai dengan batas toleransi risiko mereka. Saham yang termasuk dalam cluster 1 atau kategori risiko rendah biasanya berasal dari perusahaan dengan kesehatan keuangan yang solid, memiliki penilaian yang lebih hati-hati, dan lebih sesuai untuk investor yang mengedepankan keamanan investasi jangka panjang. Sementara itu, saham dalam kelompok 2 atau kategori risiko menengah memiliki peluang pertumbuhan yang lebih tinggi, namun juga disertai dengan tingkat risiko yang lebih signifikan, sehingga lebih cocok bagi investor dengan profil risiko moderat yang ingin memaksimalkan keuntungan sambil tetap mempertimbangkan faktor risiko. Di sisi lain, saham dalam kluster 3 atau kategori risiko tinggi menunjukkan fluktuasi yang lebih besar serta lebih responsif terhadap perubahan kondisi pasar, sehingga lebih sesuai untuk investor dengan toleransi risiko tinggi yang berfokus pada strategi spekulatif atau jangka pendek. Oleh karena itu, hasil *clustering* ini bisa memberikan arahan yang lebih terorganisir bagi investor dalam membuat keputusan investasi yang lebih berinformasi dan sesuai dengan profil risiko mereka.

5. KESIMPULAN DAN SARAN

Berdasarkan hasil penelitian yang telah dilakukan mengenai Penerapan Data Mining dalam Pengelompokan Saham Melalui Rasio Keuangan pada Saham Papan Utama di Bursa Efek Indonesia Menggunakan Metode K-Means Clustering, dapat disimpulkan bahwa: Penggunaan metode *K-Means Clustering* dalam pengelompokan saham berdasarkan rasio keuangan menunjukkan bahwa saham yang terdaftar di Papan Utama Bursa Efek Indonesia (BEI) dapat dikelompokkan ke dalam tiga kategori risiko investasi. Dari hasil analisis manual, diperoleh 12 saham masuk dalam kategori risiko rendah (Cluster 1), 223 saham dalam kategori risiko sedang (Cluster 2), dan 3 saham dalam kategori risiko tinggi (Cluster 3). Perbandingan hasil *clustering* manual dengan RapidMiner menunjukkan adanya perbedaan dalam distribusi jumlah saham pada setiap klaster. RapidMiner mengelompokkan saham dengan distribusi 72 saham dalam Cluster 1 (risiko rendah), 157 saham dalam Cluster 2 (risiko sedang), dan 9 saham dalam Cluster 3 (risiko tinggi). Perbedaan ini disebabkan oleh perbedaan dalam inisialisasi centroid dan jumlah iterasi dalam masing-masing metode. Analisis terhadap variabel rasio keuangan menunjukkan bahwa *Price Earnings Ratio* (PER), *Earnings Per Share* (EPS), *Book Value Per Share* (BVPS), dan *Price to Book Value* (PBV) memiliki tingkat kontribusi yang berbeda dalam membentuk klaster saham. Hasil analisis menunjukkan bahwa saham dengan EPS tinggi dan PER rendah cenderung masuk ke dalam kategori risiko rendah, sedangkan saham dengan PBV tinggi dan PER tinggi lebih dominan dalam kategori risiko tinggi. Hasil *clustering* menunjukkan efektivitas metode *K-Means Clustering* dalam membantu investor mengelompokkan saham berdasarkan karakteristik fundamentalnya. Penggunaan metode ini memungkinkan investor untuk lebih mudah mengidentifikasi saham yang sesuai dengan profil risiko mereka, sehingga dapat mendukung pengambilan keputusan investasi yang lebih terinformasi. Validasi *clustering* menggunakan rasio BCV/WCV menghasilkan nilai 1,902, yang menunjukkan bahwa variasi antar klaster lebih tinggi dibandingkan variasi dalam klaster. Hal ini mengindikasikan bahwa metode *K-Means Clustering* cukup efektif dalam mengelompokkan saham berdasarkan kesamaan karakteristik keuangan, dengan tingkat pemisahan yang baik antar kelompok saham. Implikasi hasil penelitian ini bagi investor adalah bahwa saham dalam Cluster 1 (risiko rendah) lebih cocok bagi investor yang mencari stabilitas dan pertumbuhan jangka panjang, Cluster 2 (risiko sedang) sesuai bagi investor dengan toleransi risiko menengah yang menginginkan keseimbangan antara risiko dan profitabilitas, sedangkan Cluster 3 (risiko tinggi) lebih cocok bagi investor agresif yang siap menghadapi volatilitas harga saham yang tinggi.

Agar penelitian ini dapat lebih dikembangkan dan memberikan manfaat yang lebih luas dalam analisis investasi saham, maka beberapa saran yang dapat diberikan adalah sebagai berikut: Penambahan variabel keuangan dalam analisis clustering, seperti *Debt to Equity Ratio* (DER) dan *Return on Assets* (ROA), guna memberikan gambaran yang lebih komprehensif mengenai kinerja keuangan perusahaan. Eksplorasi metode clustering lainnya, seperti *DBSCAN* atau *Hierarchical Clustering*, untuk membandingkan hasil pengelompokan dan mengevaluasi apakah metode lain dapat menghasilkan klasifikasi yang lebih akurat dan relevan. Dengan adanya penelitian ini, diharapkan bahwa penerapan *K-Means Clustering* dapat menjadi salah satu metode yang bermanfaat bagi investor dalam mengelompokkan saham berdasarkan rasio keuangan, sehingga dapat meningkatkan efektivitas dalam pengambilan keputusan investasi dan mengurangi risiko kerugian di pasar modal.

DAFTAR PUSTAKA

- [1] Mulyono, "ANALISA KORELASI RETURN INDEKS-INDEKS SAHAM TERHADAP INDEKS HARGA SAHAM GABUNGAN PADA BURSA EFEK INDONESIA," 2021.
- [2] P. Chang Hartono and A. Dwiyo Widianoro, "Analisis Prediksi Harga Saham Unilever Menggunakan Regresi Linier dengan RapidMiner," 2024. [Online]. Available: <https://journal-computing.org/index.php/journal-cisa/index>
- [3] P. Hanjasto, "Analisis Trend Saham LQ-45 selama Pandemi Covid-19 dengan K-Means Clustering," *Jurnal Informatika Ekonomi Bisnis*, pp. 186–191, Sep. 2022, doi: 10.37034/infec.v4i4.169.
- [4] I. Sekar, D. Hasnida, and R. Kusumawati, "Penerapan Model-Based Clustering ... / PENERAPAN MODEL-BASED CLUSTERING PADA PENGELOMPOKAN SAHAM BERDASARKAN RASIO KEUANGAN," 2023. [Online]. Available: www.yahoofinance.com
- [5] I. Menggunakan, M. Sharpe, D. Jensen, A. Karim, A. Hamid, and I. F. Cahyadi, "Analisis Kinerja Reksadana Saham Syariah Di Pasar Modal," *MALIA: Journal of Islamic Banking and Finance*, vol. 3, no. 2, 2022.
- [6] V. Onasie and S. Widoatmodjo, "Niat Investasi Generasi Milenial Di Pasar Modal," *Jurnal Manajerial Dan Kewirausahaan*, vol. 2, no. 2, p. 318, May 2020, doi: 10.24912/jmk.v2i2.7924.
- [7] A. N. Fadila and C. Nuswandari, "Apa Saja Faktor-Faktor Yang Mempengaruhi Harga Saham?," 2022. [Online]. Available: <http://journal.stekom.ac.id/index.php/E-Bisnis/page283>
- [8] K. S. Hermawan and K. D. Hartomo, "InfoTekJar : Jurnal Nasional Informatika dan

- Teknologi Jaringan Attribution-NonCommercial 4.0 International. Some rights reserved Sistem Pendukung Keputusan Sistem Pendukung Keputusan Menyeleksi Saham LQ45 untuk Generasi Milenial Menggunakan Metode SAW,” vol. 5, no. 2, 2021, doi: 10.30743/infotekjar.v5i2.3455.
- [9] A. Imelda, S. Angelica Cindiyasari Sihono, and D. Riski Anggarini, “PENGARUH LIKUIDITAS, PROFITABILITAS, DAN RASIO PASAR TERHADAP HARGA SAHAM (STUDI KASUS PADA PERUSAHAAN INDEKS LQ45 YANG TERDAFTAR DI BURSA EFEK INDONESIA PERIODE 2017-2021),” 2022. [Online]. Available: <http://jurnal.teknokrat.ac.id/index.php/JEB>
- [10] D. Tohendry and D. Jollyta, “PENERAPAN ALGORITMA K-MEANS CLUSTERING UNTUK PENGELOMPOKKAN SAHAM BERDASARKAN PRICE EARNING RATIO DAN PRICE TO BOOK VALUE,” *Jurnal Mahasiswa Aplikasi Teknologi Komputer dan Informasi*, vol. 5, no. 1, 2023.
- [11] Intan Kurniati and Aria Aji Priyanto, “Pengaruh Earning Per Share (EPS) dan Book Value Per Share (BVPS) Terhadap Harga Saham PT Telekomunikasi Indonesia TBK Periode 2012-2021,” *Formosa Journal of Sustainable Research*, vol. 1, no. 3, pp. 425–440, Aug. 2022, doi: 10.55927/fjsr.v1i3.910.
- [12] M. M. Bode, S. Murni, and F. V. Arie, “ANALISIS PRICE EARNING RATIO, PRICE TO BOOK VALUE, RETURN ON EQUITY, RISIKO TERHADAP HARGA SAHAM LQ45 PERUSAHAAN KONSTRUKSI DAN PROPERTI DI BURSA EFEK INDONESIA ANALYSIS OF PRICE EARNING RATIO, PRICE TO BOOK VALUE, RETURN ON EQUITY, AND RISK ON LQ45 INDEX,” 2022.
- [13] V. R. Prasetyo, H. Lazuardi, A. A. Mulyono, and C. Lauw, “Penerapan Aplikasi RapidMiner Untuk Prediksi Nilai Tukar Rupiah Terhadap US Dollar Dengan Metode Linear Regression,” *Jurnal Nasional Teknologi dan Sistem Informasi*, vol. 7, no. 1, pp. 8–17, May 2021, doi: 10.25077/tekno.v7i1.2021.8-17.