BAB III
METODE PENELITIAN

3.1 Objek Penelitian

Menurut Surokim (2016), objek penelitian adalah sifat keadaan dari suatu benda, orang, atau keadaan yang menjadi pusat perhatian atau sasaran penelitian. Berdasarkan definisi tersebut, objek dalam penelitian ini adalah data kinerja keuangan perusahaan sub-sektor perbankan yang terdaftar di Bursa Efek Indonesia (BEI) selama periode 2021-2023. Secara spesifik, ciri atau karakteristik yang menjadi fokus utama mencakup tiga dimensi fundamental, yaitu profitabilitas yang diukur menggunakan *Net Interest Margin* (NIM) dan *Return on Assets* (ROA); risiko kredit yang diukur menggunakan *Non-Performing Loan* (NPL), *Loan to Deposit Ratio* (LDR), dan Rasio Cadangan Kecukupan Penurunan Nilai (CKPN) terhadap Aset Produktif; serta Valuasi saham yang diukur dengan *Price-to-Book Value* (PBV) dan *Price to Earning Ratio* (PER). Penelitian ini tidak hanya terbatas pada deskripsi masing-masing variabel secara terpisah, tetapi lebih mendalam pada proses untuk mengeksplorasi pola interaksi di antara ketiganya guna membentuk klaster-klaster yang bermakna.

3.2 Desain Penelitian

Penelitian ini merupakan studi deskriptif kuantitatif yang menggunakan data sekunder berupa laporan keuangan perusahaan yang terdaftar di Bursa Efek Indonesia (BEI) tahun 2021-2023. Penelitian eksploratif merupakan penelitian awal yang bertujuan untuk mendapatkan gambaran mengenai suatu topik penelitian untuk nantinya akan diteliti lebih jauh (Morrisan, 2017). Dalam penelitian ini, pendekatan eksploratif digunakan untuk memahami karakteristik perusahaan berdasarkan kinerja keuangan dengan memanfaatkan algoritma *K-Means Clustering*.

Pendekatan kuantitatif dipilih karena penelitian ini melibatkan analisis data numerik dan statistik dari laporan keuangan perusahaan. Data sekunder yang digunakan diperoleh dari sumber resmi seperti laporan tahunan dan laporan keuangan yang tersedia di situs Bursa Efek Indonesia serta publikasi lainnya. Data tersebut kemudian diolah untuk mengekstraksi informasi keuangan yang relevan dengan pengelompokan saham menggunakan algoritma *K-Means Clustering*.

3.3 Definisi dan Operasionalisasi Variabel

Dalam penelitian ini, variabel independen merujuk pada rangkaian indikator fundamental yang digunakan sebagai dasar untuk melakukan eksplorasi dan pengelompokan perusahaan perbankan di Bursa Efek Indonesia. Meskipun dalam analisis klaster tidak ada hubungan sebab-akibat layaknya penelitian kausal, variabel-variabel ini berfungsi sebagai input utama yang akan membentuk karakteristik unik dari setiap klaster yang dihasilkan (Hair et al., 2019; Han et al., 2023). Penelitian ini secara spesifik berfokus pada tiga dimensi fundamental, yakni profitabilitas, risiko kredit, dan valuasi saham, yang secara kolektif membentuk potret kesehatan finansial dan persepsi pasar terhadap sebuah bank.

Pemilihan ketiga variabel ini didasarkan pada relevansinya dalam analisis fundamental. Dimensi profitabilitas mengukur kemampuan perusahaan menghasilkan laba dari aktivitas operasional utamanya. Risiko kredit mengidentifikasi eksposur risiko paling signifikan dalam industri perbankan. Valuasi saham mencerminkan sentimen serta ekspektasi investor. Dengan menjadikan representasi kuantitatif dari ketiga dimensi ini sebagai input, algoritma *K-Means* dapat diimplementasikan untuk melakukan eksplorasi data secara efektif. Untuk memahami lebih dalam bagaimana setiap variabel ini didefinisikan secara operasional dan diukur dalam penelitian ini disajikan dalam **Tabel 3.1.**

Tabel 3.1 Operasionalisasi Variabel

Variabel	Definisi Variabel	Indikator	Skala Data
Profitabilitas	Profitabilitas perbankan adalah kemampuan bank untuk menghasilkan laba dari aktivitas operasionalnya selama periode tertentu (Alafiyah & Priharta, 2024).	*Net Interest Margin* (NIM) = $\frac{Pendapatan\ bunga - beban\ bunga}{Rata{-}rata\ aset\ produktif} \times 100$ *Return on Asset* (ROA) = $\frac{Laba\ Bersih}{Total\ Aset} \times 100$	Rasio
Risiko Kredit	Potensi kerugian yang dialami bank akibat ketidakmampuan debitur dalam memenuhi kewajiban pembayaran pinjaman yang telah diberikan (Purwoko & Sudiyatno, 2013).	*Non Performing Loan* (NPL) = $\frac{Kredit\ Bermasalah}{Total\ Kredit} \times 100$ *Loan to Debt Ratio* (LDR) = $\frac{Kredit}{Dana\ Pihak\ Ketiga} \times 100$ Rasio CKPN terhadap aset produktif = $\frac{CKPN\ Aset\ Keuangan}{Total\ Aset\ Produktif} \times 100$	Rasio

Variabel	Definisi Variabel	Indikator	Skala Data
Valuasi Saham	Metode evaluasi untuk mengukur nilai intrinsik sebuah saham dengan membandingkan harga pasar saham dengan nilai buku per saham (Damodaran, 2006).	Nilai Buku per Lembar Saham = $\frac{Total\ Ekuitas}{Jumlah\ Saham\ Beredar}$ *Price to Book Value* (PBV) = $\frac{Harga\ Pasar\ Saham}{Nilai\ Buku\ per\ Lembar\ Saham}$ *Price Earning Ratio* (PER) = $\frac{Harga\ Pasar\ Saham}{Laba\ per\ Lembar\ Saham}$	Rasio

3.4 Populasi dan Sampel Penelitian

3.4.1 Populasi Penelitian

Populasi menurut Ibrahim (2018) adalah keseluruhan obyek yang diteliti. Populasi dalam penelitian ini adalah seluruh perusahaan perbankan yang terdaftar di Bursa Efek Indonesia (BEI) tahun 2021-2023. Berdasarkan data dari laman Bursa Efek Indonesia, terdapat 47 perusahaan perbankan yang terdaftar di BEI. Populasi ini dipilih karena mencerminkan seluruh dinamika pasar modal Indonesia dan memberikan cakupan data yang luas untuk dianalisis dengan algoritma *K-Means Clustering*.

3.4.2 Sampel Penelitian

Sampel sampel adalah bagian dari populasi yang ditentukan oleh peneliti dengan mempertimbangkan beberapa hal seperti, masalah penelitian yang dihadapi, tujuan penelitian yang ingin dicapai, hipotesis penelitian, metode penelitian, dan instrumen penelitian (Surokim et al., 2016). Sampel pada penelitian ini termasuk jenis non probabilitas, dimana sampel yang dipilih sedemikian rupa dari populasi

sehingga setiap anggota tidak memiliki probabilitas yang sama untuk dijadikan sampel. Teknik yang digunakan adalah *purposive sampling*. Teknik ini sering juga disebut sebagai penentuan sampel berdasarkan pertimbangan tertentu (*judgement sampling*) (Surokim et al., 2016). Penentuan sampel dengan teknik ini dilakukan dengan cara memilih sumber data berdasarkan kriteria tertentu yang relevan dengan tujuan penelitian (Ibrahim et al., 2018).

Penelitian ini memilih untuk menggunakan sampel dari perbankan konvensional saja, karena metode penghitungan *Net Interest Margin* (NIM) pada perbankan syariah berbeda secara prinsip dari konvensional. Pada bank syariah, istilah NIM secara teknis diganti dengan *Net Investment Margin*, yang mencerminkan pendapatan dari investasi dan akad syariah, sehingga tidak sebanding dengan NIM berbasis bunga konvensional. Akibatnya, membandingkan NIM dari kedua sistem tanpa penyesuaian metodologis akan menghasilkan analisis yang tidak sebanding dan bias (Aladin & Maudos, 2013). Oleh karena itu, untuk menjaga konsistensi metodologi dan validitas perbandingan, penelitian ini hanya menggunakan perbankan konvensional sebagai sampel sebanyak 43 perusahaan seperti pada **Tabel 3.2**.

Tabel 3.2 Pemilihan Sampel Penelitian

Perusahaan sub-sektor perbankan yang terdaftar di Bursa Efek Indonesia dari tahun 2021 - 2023	47
Perusahaan perbankan syariah	(4)
Jumlah perusahaan sampel penelitian	43
Pengamatan selama 3 tahun	× 3
Jumlah titik data	129

3.5 Teknik Pengumpulan Data

Salah satu teknik pengumpulan data dapat dilakukan dengan cara dokumentasi. Dokumentasi melibatkan pengumpulan data dari dokumen, arsip, atau bahan tertulis lainnya yang berkaitan dengan fenomena penelitian (Surokim et al., 2016). Data yang digunakan dalam penelitian ini adalah data rasio-rasio keuangan perusahaan seperti: rasio profitabilitas yang diproksikan oleh Net Interest Margin (NIM), rasio risiko kredit yang diukur melalui Non-Performing Loan (NPL), dan rasio valuasi saham yang direpresentasikan oleh Price to Book Value (PBV). Data ini termasuk ke dalam jenis data sekunder yang diperoleh dari situs resmi Bursa Efek Indonesia. Selain itu, data pendukung lainnya diambil dari publikasi lembaga keuangan, platform analisis pasar modal, dan jurnal terkait. Data ini kemudian disusun dalam format tabular untuk memudahkan proses analisis lebih lanjut.

3.6 Teknik Analisis Data

Setelah data dikumpulkan analisis statistik deskriptif digunakan untuk memberikan gambaran umum mengenai data rasio keuangan yang dianalisis. Statistik deskriptif meliputi mencari apakah ada data yang hilang, penghitungan nilai minimum, maksimum, rata-rata, dan standar deviasi dari rasio keuangan yang digunakan. Proses ini bertujuan untuk memahami distribusi data dan mengidentifikasi pola awal sebelum menerapkan algoritma K-Means Clustering (Jumairi & Purnama, 2025). Analisis ini dilakukan menggunakan perangkat perangkat lunak Orange Data Mining.

3.6.1 *Preprocessing* Data

Tahapan *pre-processing* data sangat penting untuk memastikan kualitas dan relevansi data yang dianalisis. Pre-processing meliputi serangkaian proses untuk membersihkan, mentransformasi, dan menyiapkan data mentah agar sesuai untuk analisis dan menghasilkan model yang akurat dan efisien (Han et al., 2023).

Pengumpulan data harus dilakukan terlebih dahulu. Setelah semua data yang dibutuhkan sudah terkumpul, langkah selanjutnya adalah pembersihan data. Proses ini menangani beberapa masalah potensial dalam data, antara lain nilai hilang dan inkonsistensi data. Setelah data dibersihkan, tahap selanjutnya adalah penskalaan data yangsangat penting dalam *K-Means Clustering* karena algoritma ini sensitif terhadap skala variabel. Variabel dengan skala yang lebih besar dapat mendominasi perhitungan jarak antar data (Naufal et al., 2024).

3.6.2 Penentuan Jumlah Klaster

Elbow Method adalah teknik yang digunakan untuk menentukan jumlah klaster optimal dalam algoritma klasterisasi, seperti K-Means. Metode Elbow bekerja dengan cara membandingkan persentase hasil perbandingan antara jumlah klaster yang membentuk siku pada suatu titik (Tohendry & Jollyta, 2023b). Langkah-langkah metode elbow:

1. Menginisialisasi nilai awal k
2. Meningkatkan nilai k.
3. Menghitung nilai Sum of Square Error untuk setiap nilai k.
4. Menganalisis Sum of Square Error untuk nilai k yang mengalami penurunan secara signifikan.
5. Menetapkan nilai k pada titik dimana terdapat siku pada grafik hasil analisis Sum of Square Error.

Berikut Merupakan rumus dari metode Elbow :

$$\text{SSE} = \sum_{k=1}^{k} \quad \sum_{xi \epsilon Sk} \quad ||Xi - Ck||^2$$

Keterangan:

K: Jumlah klaster

Xi: Data ke-i

Ck: Centroid klaster

Meskipun Elbow Method sederhana dan efektif, pendekatan ini memiliki keterbatasan. Salah satu tantangan utama adalah bahwa tidak semua dataset menghasilkan "sudut siku" yang jelas. Dalam kasus seperti ini, pendekatan

tambahan, seperti Silhouette Score, mungkin diperlukan untuk mendukung keputusan akhir terkait jumlah klaster optimal (Ridwan et al., 2021).

3.6.3 Penerapan Algoritma K-Means Clustering

K-Means Clustering adalah algoritma unsupervised learning yang digunakan untuk mengelompokkan data ke dalam beberapa klaster berdasarkan kesamaan karakteristik (Han et al., 2011) Perangkat lunak yang akan digunakan dalam proses klasterisasi menggunakan K-means adalah *Orange Data Mining*. Proses analisis melibatkan langkah-langkah berikut (Tohendry & Jollyta, 2023b):

1. menentukan jumlah kluster yang optimal
2. menginisialisasi centroid awal
3. mengelompokkan data berdasarkan jarak Euclidean dengan rumus sebagai berikut:

   $$D(i, j) = \sqrt{(X1i - X1j)2 + (X2i - X2j)2 + \cdots + (Xki - Xkj)2}$$

   Dimana D (i,j) = Jarak data ke i ke pusat cluster j

   Xki = Data ke i pada atribut data ke k

   Xkj = Titik pusat ke j pada atribut ke k
4. memperbarui centroid hingga mencapai stabilitas kluster atau disebut juga iterasi.

Hasil dari analisis ini akan dianalisis sehingga menggambarkan karakteristik saham dalam masing-masing kluster dan memberikan wawasan yang relevan bagi investor untuk pengambilan keputusan investasi.

3.6.4 Validasi Hasil Klastering

Silhouette Method adalah teknik yang digunakan untuk mengevaluasi kualitas klasterisasi dengan menilai seberapa baik data dalam klaster tertentu cocok dengan klaster mereka sendiri dibandingkan dengan klaster lain. Nilai silhouette mengukur koherensi suatu data dengan klaster tempat ia berada serta perbedaannya dengan klaster lain, memberikan pemahaman mendalam tentang struktur klaster (Rousseeuw, 1987). Berikut adalah rumus untuk menghitung Silhouette Score:

$$s(i) = \frac{b(i) - a(i)}{max\{a(i),b(i)\}}$$

Keterangan:

s(i): Nilai atau skor siluet untuk satu objek (data point) i. Nilainya berkisar dari -1 hingga +1.

a(i): Rata-rata ketidaksamaan dari objek i ke semua objek lain dalam kluster yang sama. Nilai ini mengukur seberapa baik sebuah objek ditempatkan dalam klusternya (kohesi). Semakin kecil nilainya, semakin baik.

b(i): Rata-rata ketidaksamaan terkecil dari objek i ke semua objek dalam kluster lain (yaitu, kluster tetangga terdekat). Nilai ini mengukur seberapa jauh sebuah objek dari kluster lain (separasi).

Nilai s(i) berkisar antara -1 dan 1 (Rousseeuw, 1987). Nilai mendekati 1 menunjukkan bahwa data sangat cocok dengan klaster-nya dan tidak cocok dengan klaster lain. Nilai mendekati 0 menunjukkan bahwa data berada di perbatasan antara dua klaster. Nilai negatif menunjukkan bahwa data lebih cocok berada di klaster lain dibandingkan klaster saat ini.

3.6.5 Interpretasi Klaster Hasil

Tahap interpretasi klaster mencakup pemeriksaan setiap kelompok dari segi variasi klaster dengan cara memberikan nama atau menandai dengan suatu label secara tepat sehingga dapat menggambarkan sifat dari suatu klaster tersebut (Syukria et al., 2022). Interpretasi klaster hasil bertujuan untuk memahami karakteristik dan makna dari kelompok-kelompok yang terbentuk setelah proses klasterisasi. Selain itu, pemetaan variabel dominan dalam klaster dapat

memberikan wawasan lebih lanjut mengenai faktor utama yang membedakan satu klaster dengan lainnya. Untuk memahami lebih dalam setiap klaster, analisis statistik deskriptif dapat digunakan guna mengevaluasi distribusi data dalam masing-masing kelompok. Penggunaan ukuran statistik seperti rata – rata dan standar deviasi dapat membantu dalam mengidentifikasi tren dan perbedaan signifikan antara klaster (Jumairi & Purnama, 2025; Syukria et al., 2022; Wisna et al., 2023). Mean memberikan gambaran umum mengenai nilai rata-rata dalam suatu klaster. Standar deviasi, di sisi lain, menunjukkan seberapa besar variasi data dalam klaster, yang dapat membantu menentukan apakah klaster bersifat homogen atau heterogen.