

BAB II

TINJAUAN PUSTAKA

2.1 Investasi

Investasi adalah komitmen atas sejumlah dana atau sumber daya lainnya yang dilakukan pada saat ini, dengan tujuan memperoleh sejumlah keuntungan dimasa datang (Tendelilin, 2001). Istilah investasi bisa berkaitan dengan berbagai macam aktivitas. Menginvestasikan dana pada sektor rill (tanah, emas, mesin atau bangunan) maupun asset finansial (deposito, saham atau obligasi) merupakan aktifitas yang umum di lakukan.

Menurut Jogiyanto, investasi dapat didefinisikan sebagai penundaan konsumsi sekarang untuk digunakan dalam produksi yang efisien selama periode waktu tertentu (Jogiyanto, 2003). Sedangkan menurut Sukirno kegiatan investasi yang dilakukan oleh masyarakat secara terus menerus akan meningkatkan kegiatan ekonomi dan kesempatan kerja, meningkatkan pendapatan nasional dan meningkatkan taraf kemakmuran masyarakat. Peranan ini bersumber dari tiga fungsi penting dari kegiatan investasi, yakni (1) investasi merupakan salah satu komponen dari pengeluaran agregat, sehingga kenaikan investasi akan meningkatkan permintaan agregat, pendapatan nasional serta kesempatan kerja (2) penambahan barang modal sebagai akibat investasi akan menambah kapasitas produksi (3) investasi selalu diikuti oleh perkembangan teknologi (Nizar, et.al, 2013).

2.2 Pasar Modal

Pasar modal merupakan alternatif bagi mahasiswa atau investor untuk berinvestasi baik dalam jangka pendek maupun jangka panjang. Saham ialah salah satu produk keuangan pasar modal yang diperjual belikan dan yang paling populer. Saham merupakan surat berharga sebagai tanda bukti kepemilikan atau penyertaan dari perorangan maupun instansi dalam suatu perusahaan (Anoraga dan Pakarti, 2006:58). Sebelum melakukan investasi pada satu instrumen investasi, Investor mempelajari segala hal yang berhubungan dengan investasi. Para investor menyadari bahwa investasi dapat menguntungkan dimasa depan.

2.3 Saham

Saham merupakan surat bukti kepemilikan atas aset-aset perusahaan yang menerbitkan saham. Saham dapat dibedakan menjadi saham preferen dan saham biasa. Harga saham adalah nilai suatu saham yang mencerminkan kekayaan perusahaan yang mengeluarkan saham tersebut, dimana perubahan atau fluktuasinya sangat ditentukan oleh kekuatan permintaan dan penawaran yang terjadi di pasar bursa (pasar sekunder). Semakin banyak investor yang ingin membeli atau menyimpan saham, harganya semakin naik. Sebaliknya semakin banyak investor yang ingin menjual atau melepaskan suatu saham, maka harganya semakin bergerak turun.

2.4 Evaluasi Instrumen Pasar Modal

Yang dimaksud instrumen pasar modal adalah semua surat-surat yang berharga yang diperdagangkan di bursa. Instrumen pasar modal ini umumnya bersifat jangka panjang.

Evaluasi keadaan ekonomi secara makro dan tren industri memegang peranan penting dalam investasi di Pasar Modal. Kebijakan moneter dan kebijakan fiskal yang dikeluarkan oleh pemerintah, seperti peraturan perpajakan, kebijakan tingkat suku bunga, belanja pemerintah, serta tingkat inflasi yang terjadi akan memberikan pengaruh yang signifikan ke seluruh industri, dan perusahaan/emiten. Kebijakan-kebijakan tersebut dapat memberikan pengaruh positif, negative ataupun netral terhadap pendapatan, pengeluaran dan laba perusahaan/emiten, yang pada akhirnya akan mempengaruhi tingkat pengembalian investasi.

2.5 Data Mining

Menurut Gartner Group data mining adalah proses menemukan hubungan baru yang mempunyai arti, pola dan kebiasaan dengan memilah-milah sebagian besar data yang disimpan dalam media penyimpanan dengan menggunakan teknologi pengenalan pola seperti teknik statistik dan matematika. Data mining merupakan gabungan dari beberapa disiplin ilmu yang menyatukan teknik dari pembelajaran mesin, pengenalan pola, statistik, database, dan visualisasi untuk penanganan permasalahan pengambilan informasi dari database yang besar.

Data mining menurut David Hand, Heikki Mannila, dan Padhraic Smyth dari MIT adalah analisa terhadap data (biasanya data yang berukuran besar) untuk menemukan hubungan yang jelas serta menyimpulkannya yang belum diketahui sebelumnya dengan cara terkini dipahami dan berguna bagi pemilik data tersebut.

Menurut Sumathi dan Sivandham (2006), data mining juga didefinisikan sebagai bagian dari proses penggalian pengetahuan dalam data base yang sering disebut dengan istilah *Knowledge Discovery in Database* (KDD). KDD merupakan

suatu area yang mengintegrasikan berbagai metode, yang meliputi statistik, basis data, kecerdasan buatan (*Artificial Intelligence*), *machine learning*, pengenalan pola (*Pattern Recognition*), pemodelan yang menangani ketidak pastian, visualisasi data, optimasi, Sistem Informasi Manajemen (SIM), dan sistem berbasis pengetahuan (*knowledge based-system*).

2.6 Statistika Deskriptif

Statistika deskriptif adalah metode statistika yang berfungsi untuk memberikan gambaran umum tentang penyajian sampel dan populasi. Analisis statistika deskriptif dapat diartikan sebagai metode yang berkaitan dengan mengumpulkan, mengolah, dan menyajikan data sehingga memberikan informasi yang berguna (Walpole, 1998). Data dapat disajikan dalam bentuk grafik atau tabel, untuk ukuran pemusatan data dan ukuran penyebaran data dideskripsikan secara numerik. Ukuran pemusatan data meliputi rata-rata, nilai tengah, dan modus sedangkan ukuran penyebaran data meliputi ragam dan standar deviasi menurut (Walpole dkk. 2012).

2.7 Pengujian Distribusi Normal Multivariat

Melalui Chi-square plot, kita bias melakukan uji asumsi multivariate normal:

1. Jika plot $\{qc, p \frac{j-1}{n}, d_j^2\}$ pada Chi-square plot mengikuti fit line, maka asumsi multivariate normal terpenuhi
2. Jika korelasi antara $\{qc, p \frac{j-1}{n}, d_j^2\}$ memiliki nilai yang besar, maka asumsi multivariat normal terpenuhi

Adapun langkah-langkah dalam menggambar Chi-square plot, yaitu:

1. Menghitung kuadrat jarak mahalobis $d^2 = (x_j - \bar{x})' S^{-1}(x_j - \bar{x})$
2. Urutkan kuadrat jarak dari yang terkecil ke yang terbesar $d_1^2, d_2^2, \dots, d_p^2$
3. Menghitung nilai kuantil untuk chi-square distribution dengan $df = p$

$$q_{c,p} \frac{j-\frac{1}{2}}{n} : \frac{100(j-\frac{1}{2})}{n} \text{ dimana } q_{c,p} \frac{j-\frac{1}{2}}{n} X_p^2 (1 - \frac{j-\frac{1}{2}}{n})$$

4. Gambar Chi-square Plot dengan nilai $\{q_{c,p} \frac{j-\frac{1}{2}}{n}, d_j^2\}$

2.8 Pengujian Kecukupan Data

Measurement of Sampling Adequacy (MSA) adalah indeks perbandingan jarak antara koefisien korelasi dengan koefisien korelasi parsialnya. Berdasarkan Hair dkk. (2009), apabila kuadrat dari elemen matriks korelasi D di antara seluruh pasangan variabel bernilai lebih kecil jika dibandingkan dengan jumlah kuadrat dari elemen matriks korelasi R, maka akan menghasilkan nilai MSA mendekati 1.

Nilai MSA dianggap mencukupi jika lebih dari 0,5. Analisis MSA berfungsi untuk menguji kecukupan data yang telah terambil meliputi segi kecukupan data untuk difaktorkan. Rumus yang digunakan ditunjukkan pada persamaan.

$$MSA = \frac{\sum_{i=1}^p \sum_{j=1}^p r_{ij}^2}{\sum_{i=1}^p \sum_{j=1}^p r_{ij}^2 + \sum_{i=1}^p \sum_{j=1}^p q_{ij}^2}$$

Dimana:

r_{ij}^2 = kuadrat dari elemen matriks korelasi **R**

q_{ij}^2 = kuadrat dari elemen matriks korelasi D = $R^{-1}D$

$i, j = 1, 2, 3, \dots, p$

2.9 Pengujian Independensi

Uji independensi digunakan untuk mengetahui besarnya nilai korelasi antar variabel. Apabila variabel x_1, x_2, \dots, x_N independen (bersifat saling bebas), maka matriks korelasi antar variabel sama dengan matriks identitas (Rencher, 1998). Berikut ini adalah hipotesis yang digunakan dalam *Bartlett Test*.

$H_0 : \rho = I$ (tidak terdapat korelasi)

$H_1 : \rho \neq I$ (terdapat korelasi)

Dengan statistik uji ditunjukkan pada Persamaan sebagai berikut:

$$X^2 = -\left(C - 1 - \frac{2N + 5}{6}\right) \ln|R|$$

Dimana:

$|R|$ = nilai determinan dari matriks korelasi

N = banyak variabel

Adapun H_0 akan ditolak jika $X^2 > X^2_{\alpha, \frac{1}{2}N(N-1)}$ yang berarti terdapat korelasi multivariat antar variabel penelitian (Rencher, 1998).

2.10 Penentuan Jumlah Faktor

Penentuan jumlah faktor yang dibentuk dapat menggunakan tiga kriteria yang biasa digunakan dalam analisis komponen utama (Rencher, 1998). Kriteria yang paling sering digunakan pada analisis komponen utama adalah penentuan banyak faktor berdasarkan proporsi faktor ke- j dalam menjelaskan total varians sampel. Matriks korelasi yang bernilai lebih dari 1 dapat menjadi faktor berdasarkan kriteria *Eigenvalues*. Selain penentuan banyak faktor dapat dilihat secara visual melalui *scree plot*.

2.10.1 Rotasi Faktor

Setelah diketahui jumlah faktor yang terbentuk, tahapan selanjutnya adalah menginterpretasikan komponen dari faktor tersebut. Terkadang loading faktor yang dihasilkan memiliki nilai yang hampir sama pada setiap kolomnya, sehingga akan terjadi kesulitan dalam menginterpretasikannya. Rotasi faktor bertujuan untuk menyederhanakan struktur faktor sehingga mudah untuk diinterpretasikan. Ada dua metode rotasi yang berbeda yaitu orthogonal dan oblique rotation (Rencher, 1998). Metode rotasi yang populer ialah *varimax procedure*. Prosedur ini merupakan metode orthogonal yang berusaha memaksimalkan nilai kuadrat dari loading.

2.11 Analisis Klaster

Analisis Klaster merupakan sebuah teknik untuk mengelompokkan objek ke dalam kelompok-kelompok sesuai dengan karakteristik tertentu (Johnson & Wichern, 2007). Setiap pengamatan harus memiliki homogenitas yang tinggi dalam sebuah kelompok dan memiliki heterogenitas yang tinggi dengan kelompok yang lainnya. Analisis ini akan mengelompokkan objek sehingga setiap objek yang memiliki kesamaan dengan objek lainnya akan berada dalam klaster yang sama (Sharma, 1996). Cluster analysis bertujuan untuk menentukan suatu kelompok yang alami dari kelompok-kelompok individu. Kelompok individu ini bisa membentuk populasi yang lengkap atau suatu sampel dari populasi yang lebih besar. Cluster analysis akan mengalokasikan sekelompok individu pada suatu kelompok yang saling bebas sehingga individu-individu di dalam kelompok itu mirip satu dengan yang lainnya, sementara itu individu-individu di dalam kelompok yang berbeda tidak mirip. Penyusunan kelompok ini biasa disebut dengan partisi (Ruswandi,

2008). Salah satu ukuran kemiripan yang dapat digunakan adalah jarak euclidean dan mahalanobis. Pada sudut pandang lainnya, beberapa metode cluster dapat dikelompokkan berdasarkan algoritma proses yang dilakukan, yaitu teknik yang berdasarkan ukuran jarak sebagai basis pengelompokannya.

Metode berbasis ukuran jarak ini terdiri dari metode cluster berhierarki dengan penggabungan (agglomerative) dan pemisahan (decisive). Contohnya adalah metode pautan lengkap complete linkage, metode pautan rata-rata average linkage, metode Ward serta metode cluster tak berhierarki yaitu metode k-means (Andesberg, 1973). Hierarchical clustering maupun k-means clustering hanya memperhatikan ukuran jarak antar objek pengamatan tanpa mempertimbangkan aspek statistik lainnya, seperti distribusi data ataupun objek pengamatan pada cluster yang saling tumpang tindih. Sedangkan algoritma fuzzy clustering membentuk data kedalam struktur klaster yang berbentuk spherical, ellipsoid atau bentuk yang lain.

2.12 Algoritma Clustering

2.12.1 Metode Fuzzy C-means Clustering

Metode klaster *k-means* merupakan salah satu metode klaster non hirarki dengan jumlah klaster sudah ditentukan sebelum pengelompokan yang dikembangkan (Bezdek, 1981). *Fuzzy c-means* merupakan salah satu metode pengelompokan yang dikembangkan dari *c-means* dengan menerapkan sifat *fuzzy* dalam fungsi keanggotaannya. Dalam metode ini dipergunakan variabel *membership function* yang merujuk pada seberapa besar kemungkinan suatu data

bisa menjadi anggota ke dalam suatu kelompok. *Membership function* yang digunakan dapat berasal dari skor faktor.

Berdasarkan (Balasko, et.al, 2007), langkah algoritma dari metode FCM sebagai berikut:

1. Memasukkan data yang dikelompokkan
2. Menentukan banyak jumlah kelompok yang akan dibentuk ($1 < c < N$), *weighting exponent* ($m > 1$), maksimum iterasi error terkecil yang diharapkan ($\varepsilon > 0$), fungsi objektif awal = 0, dan iterasi awal ($t = 1$)
3. Membangkitkan bilangan random $\mu_{ik}, i = 1, 2, \dots, c; 1 \leq k \leq N$ sebagai elemen matrikpartisi awal U_0

$$\sum_{i=1}^c \mu_{ik} = 1, i = 1, 2, \dots, c; 1 \leq k \leq N$$

4. Menghitung pusat kelompok ke- k : v menurut persamaan sebagai berikut:

$$v = \frac{\sum_{k=1}^N \mu_{ik}^m X_k}{\sum_{k=1}^N \mu_{ik}^m}, 1, 2, \dots, c$$

5. Menghitung jarak D_{ik}^2 dengan rumus sebagai berikut:

$$D_{ik}^2 = \|x_k - v_i\| = (x_k - v_i)^T A(x_k - v_i)$$

6. Menghitung fungsi objektif pada iterasi ke- t dengan rumus

$$J_{FCM} = (X, U, V) = \sum_{i=k}^c \sum_{k=1}^N (\mu_{ik})^m D_{ik}^k(x_k, v_i)$$

7. Menghitung nilai fungsi keanggotaan yang baru U_{t+1} sesuai persamaan sebagai berikut:

$$\mu_{ik} = \left[\sum_{j=1}^c \left(\frac{D(x_k, v_i)}{D(x_k, v_j)} \right)^{\frac{2}{m-1}} \right]^{-1}$$

8. Membandingkan nilai keanggotaan dalam matriks U hingga konvergen ketika $\|U_{t+1} - U_t\| < \varepsilon$ apabila $\|U_{t+1} - U_t\| \geq \varepsilon$ maka kembali ke langkah 4

2.12.2 Metode Gustafson Kessel Clustering

Berdasarkan Balasko dkk. (2007), pengelompokan fuzzy Gustafson-Kessel merupakan pengembangan dari fuzzy c-means (FCM). Nilai pembentuk matriks pada metode pengelompokan ini disebut adaptive distance norm yang diperbarui di setiap iterasi. Sehingga, pengelompokan ini mampu lebih menyesuaikan bentuk geometris fungsi keanggotaan yang tepat untuk sebuah himpunan data.

Analisis kluster fuzzy menitik beratkan pada pengelompokan yang memper timbangkan tingkat keanggotaan yang mencakup himpunan fuzzy sebagai landasan pembobotan. Perbedaan utama dari Gustafson Kessel dibandingkan dengan metode kluster hierarki maupun non hierarki ialah kemampuannya dalam menangani ketidakpastian. (Mauliyadi dkk. 2003) menyatakan nilai akurasi analisis kluster Gustafson Kessel pun lebih tinggi daripada analisis kluster fuzzy c-means. (Amalia dkk. 2016) memperoleh nilai icd-rate GK lebih rendah daripada FCM. Berdasarkan penelitian sebelumnya maka metode GK lebih baik digunakan dibandingkan metode FCM.

Algoritma Gustafson-Kessel mengubah fungsi perhitungan jarak menjadi fungsi jarak adaptif (*adaptive distance norm*) yang selalu diperbaharui pada setiap iterasi dengan menggunakan matriks fuzzy covariance. Algoritma Gustafson-

Kessel menggunakan fungsi jarak mahalnobis sehingga lebih dapat menyesuaikan bentuk geometris untuk sebuah himpunan data, tidak seperti Fuzzy C-Means yang mengasumsikan bahwa bentuk geometris suatu cluster adalah bulat sempurna. Meskipun Gustafson-Kessel lebih unggul dari algoritma Fuzzy C-Means, masih terdapat masalah saat matriks fuzzy covariance dari data merupakan matriks singular maka perhitungan matriks A_k tidak dapat diterapkan.

Algoritma Gustafson-Kessel Clustering secara lengkap adalah sebagai berikut: Input data yang akan dikelompokkan sebagai X (matriks $n \times p$), tentukan jumlah cluster yang akan dibentuk ($c \geq 2$), *weighting exponent* ($m > 1$), maksimum iterasi (t_{max}), error terkecil yang diharapkan (ϵ), nilai threshold (β), dan parameter pembobot $\gamma \in [0,1]$. Bangkitkan bilangan random $\mu_{ik}, 1 \leq i \leq n; 1 \leq k \leq c$ sebagai elemen-elemen matriks partisi awal U_0 dan hitung matriks kovarian F_0 dari keseluruhan data.

$$\sum_{k=1}^c \mu_{ik}^m = 1, 1 \leq i \leq n; 1 \leq k \leq c$$

Ulangi untuk $t = 1, 2, 3, \dots, t_{max}$

1. Menghitung pusat cluster ke-k (v_k) dengan rumus:

$$v_k^{(t)} = \frac{\sum_{i=1}^n (\mu_{ik}^{(t-1)})^m x_i}{\sum_{i=1}^n (\mu_{ik}^{(t-1)})}, 1 \leq k \leq c$$

Dengan:

μ_{ik} = derajat keanggotaan data ke-i pada cluster ke-k.

m = pangkat pembobot untuk fungsi keanggotaan fuzzy

t = banyaknya iterasi

n = banyaknya data

c = banyaknya cluster

2. Menghitung matriks kovarian cluster dengan rumus:

$$F_k = \frac{\sum_{i=1}^n (\mu_{ik}^{(t-1)})^m (x_i - v_k^{(t)})(x_i - v_k^{(t)})^T}{\sum_{i=1}^n (\mu_{ik}^{(t-1)})^m}, 1 \leq k \leq c$$

Dengan:

x_i = vektor data ke- i

v_k = pusat *cluster* ke- k

μ_{ik} = derajat keanggotaan data ke- i pada cluster ke- k

m = pangkat pembobot untuk fungsi keanggotaan fuzzy

t = banyaknya iterasi

n = banyaknya data

c = banyaknya cluster

Ekstraksi nilai *eigenvectors* ϕ dan *eigenvalues* λ dari F_k^{new} yang sudah dihitung dengan persamaan:

$$F_k^{new} = (1 - \gamma)F_k + \gamma \det(F_0)^{1/p} I$$

Dengan:

γ = parameter untuk mengatur bentuk matriks *fuzzy covariance*, $\gamma \in [0,1]$

F_0 = matriks kovarian dari seluruh data

F_k = matriks fuzzy covariance cluster ke- k (pada persamaan 4)

P = banyaknya variabel

I = matriks identitas

Jika rasio antara nilai eigen maksimal dan minimal melewati nilai threshold yang ditentukan, maka rekonstruksi F_k dengan penjabaran sebagai berikut:

$$F_k = \Phi \Lambda \Phi^{-1}$$

Dengan:

Φ = vektor eigen dari matriks fuzzy covariance cluster ke-k

Λ = matriks diagonal dari nilai-nilai eigen (*eigenvalues*) matriks fuzzy covariance cluster ke-k

3. Menghitung jarak dengan persamaan menggunakan persamaan sebagai berikut dengan $t=1,2,\dots,n$ dan $=1,2,\dots,c$:

$$D_{ikAk}^2 = (x_i - v_k^{(t)})^T [\rho_k \det(F_k)^{\frac{1}{p}} F_k^{-1}] (x_i - v_k^{(t)})$$

Dengan:

D_{ikAk}^2 = jarak data ke-i terhadap pusat *cluster* ke-k dengan *norm inducing matrix Ak*

x_i = vektor data ke-i

v_k = pusat cluster ke-k

F_k = matriks fuzzy covarian cluster ke-k

ρ_k = volume cluster ke-k

p = banyaknya variabel

n = banyaknya data

c = banyaknya cluster

4. Mempernaharui matriks fungsi keanggotaan untuk $1 \leq i \leq n$ jika $D_{ikAk}^2 > 0$ untuk $1 \leq k \leq c$

$$\mu_{ik}^{(t)} = [\sum_{i=1}^c (\frac{D_{ikAk}}{D_{iIAk}})^{\frac{2}{(m-1)}}]^{-1}$$

Jika tidak, maka:

$$\mu_{ik}^{(t)} = 0 \text{ jika } D_{ikAk}^2 > 0 \text{ dan } \mu_{ik}^{(t)} \in [0,1]$$

$$\text{Dengan } \sum_{k=1}^c \mu_{ik}^{(t)} = 1$$

Sampai $\|U^t - U^{(t-1)}\| < \varepsilon$ atau jika $t >$ iterasi maksimum

Nilai threshold (β) yang digunakan biasanya akan ditentukan dalam angka yang besar, seperti 10^{15} . Penentuan nilai parameter pembobot γ tergantung kepada data yang digunakan, beberapa percobaan mungkin perlu dilakukan untuk menentukan nilai γ yang tepat.

2.13 Modified Partition Coefficient Index (MPCI)

(Bezdek, 1981) mengusulkan Partition Coefficient Index (PCI) untuk evaluasi kualitas dari kluster dan menentukan secara baik data yang diwakili oleh kluster tersebut.

$$PCI = \frac{1}{n} \left(\sum_{j=1}^n \sum_{i=1}^c u_{i,j}^2 \right)$$

Akan tetapi PCI cenderung mengalami perubahan monoton terhadap nilai c yang berakibat buruknya performa dari indeks tersebut (Xie et al., 2011). Untuk mengatasi masalah tersebut (Dave, 1996) mengusulkan *Modified Partition Coefficient Index* (MPCI) untuk menghitung koefisien partisi sebagai evaluasi nilai keanggotaan data pada setiap kluster. Kluster optimal akan terbentuk ketika nilai yang diperoleh mendekati nol.

$$MPCI = 1 - \frac{C}{C-1}(1 - PCI)$$

