ISSN 1978-0524 (print) | 2528-6374 (online)
http://journal.uad.ac.id/index.php/JIFO/index

Classifying the characteristics of insurance shares: a *k-means* clustering approach

Y Utami[1], I Zuhroh[2], V Prasetya[3], M Rofik[4]

[1]Departement of Management, Universitas Pancasakti Tegal, 52121, Indonesia
[2]Department of Economics Development, Universitas Muhammadiyah Malang, 65145, Indonesia
[3]Department of Management, STIE Assholeh Pemalang, 52313, Indonesia
[4]Center for Economics, Business and Entrepreneurship, Universitas Muhammadiyah Malang, 65145, Indonesia
[1]mochamadrofik81@gmail.com*
* Corresponding Author

Received 16 February 2021; accepted August 2021; published September 2021

ABSTRACT

This study aims to apply the k-means clustering method in understanding the characteristics of insurance shares. The eight issuers are divided into three clusters based on priceand rate of return. The k-means method's application shows that each cluster has different characteristics, especially for the price variable. Test with panel data regression also discovers different patterns between clusters 2 and 3 in responding to changes in interest rates. The findings of this study indicate that k-means clustering can be used as an initial analysis to understand thecharacteristics of issuers that investors can use to increase the optimal probability of return.

KEYWORDS
k-means
insurance share
price
rate of return

1. Introduction

In General, the principle of investing is looking for a portfolio with the highest possible return and the lowest risk [1][2]. Therefore, it is important to understand the characteristics of the portfolio before making an investment decision. Understanding the characteristics of the portfolio to be purchased will greatly help investors predict the movement of returns based on the movement of predetermined exogenous variables [3]. The literature also shows that stock price movements' velocity is influenced by market capitalization, transaction volume, and the number of investors [3][4][5][6]. Therefore, each stock group has certain characteristics and patterns in responding to changes in exogenous variables such as changes in macroeconomic variables.

Data mining is an algorithm-based method utilized to discover patterns and characteristics from a set of information. A few strategies are commonly used to bunch information from a set of data; one ofthem is k-means clustering. k-means clustering is a calculation that requires as much input k, which separates n objects into k clusters. The level of likeness between individuals in each cluster is different. The nearness of centroid value measures the degree of similitude between individuals in a cluster. k-means are frequently utilized since they can classify vast sums of information with adequate computing time [7][8][9].

Due to the needs of investors in understanding specific stock price movement patterns to maximize return, opportunities to take advantage of the clustering approach, and based on our knowledge, there isno particular paper discussing the *k-means* for the Indonesian stock market, this study aims to classify the characteristics of insurance stocks listed on Indonesia Stock Exchange (IDX) using the *k-means* clustering method as a sample. Besides interpreting the cluster results descriptively, this study also conducted a regression test for each cluster by making the interest rate and exchange rate as exogenousvariables. The results of this regression can be used as additional evidence that *k-means* clustering is still powerful to be used as an initial analysis to understand a pattern from a set of observational data.

10.26555/jifo.v15i3.a23372

2. Method

Formally the *k-means* clustering method can be presented as follows. Assuming there is a set of observations (x_1, $x_{2,}x_3$ … , x_n) where each observation is a real vector with d-dimension, *k-means* clustering aims to divide n observations into k cluster ($k \leq n$). The set $S = \{S_1, S_2, S_3, \dots , S_k\}$ thus minimizing the within-cluster sum of square-like variants. The purpose of minimizing this variant can be stated as equation (1).

$$\frac{\arg min}{S}\sum_{i=1}^{k}\sum_{x\in S_i}\|x-\mu\|^2 = \frac{\arg min}{S}\sum_{i=1}^{k}|S_i| Var\ S_i \qquad (1)$$

Where μ_i is the mean of points in S_i. This is equivalent to minimizing the paired square deviations ofpoints in the same cluster as equation (2).

$$\frac{\arg min}{S}\sum_{i=1}^{k}\frac{1}{2|S_i|}\sum_{x,y\in S_i}\|x-y\|^2 \qquad (2)$$

The equivalence can be deduced from identity $\sum_{x\in S}\|x-\mu_i\|^2 = \sum_{x\neq y\in S}(x-\mu_i)(\mu_i-y)$. Since the total variance is constant, it is equivalent to maximizing the sum of the squared deviations between points in different clusters following the law of total variance.

In this paper, the total observations are eight stocks of insurance companies listed on the IDX during 2019. The data includes two-dimensional data because the observed data are monthly price and the monthly rate of return. The eight issuers are then partitioned into three clusters using the k-means algorithm using SPSS software. Each cluster that has been partitioned is regressed using panel data, withthe exogenous variables being interest rates and exchange rates. The regression results further clarify the characteristics of each cluster in responding to exogenous variables, which are commonly used as areference in predicting stock price movements. The regression results also provide additional evidence that clustering using the k-means method can be used as an initial analysis tool for understanding the specific group of issuers.

3. Results

The clustering results by dividing the issuers of insurance shares into three clusters can be seen in Table 1. Cluster 1 only has one issuer, while cluster 2 and 3 have three and four issuers, respectively. Table 2 shows that cluster 1, which is only occupied by one issuer, has an average price of IDR 62 and a returnof -1%; issuers in cluster 2 have an average price of around IDR 279 return of -2%. Meanwhile, issuersin cluster 3 have an average IDR 917 price with a return of 4%. In more detail, the prices and returns for each emitter in each cluster can be seen in Table 4. The analysis of this study also uses ANOVA as a parameter to determine whether there are really differences in prices and returns in each cluster. The ANOVA results in Table 3 show that only the price variable is significant. The results of this analysis can be interpreted that statistically, the main driving factor for cluster development in this study is the price. This is because the returns in each cluster do not really have a significant difference.

Table 1. Cluster Membership

Cluster	Issuer	Distance
1	AHAP	.000
2	AMAG	26.722
	ASBI	30.222
	ASJT	56.944
3	ASDM	135.417
	ASMI	35.000
	ASRM	35.000
	JMAS	65.417

Table 2. Final cluster center

	Cluster		
	1	2	3
Price	62.42	279.44	917.50
Return	-.01	-.02	.04

Table 3. ANOVA

	Cluster		Error		F	Sig.
	Mean Square	df	Mean Square	df		
Price	496959.494	2	5987.429	5	83.000	.000
Return	.003	2	.002	5	1.809	.256

Table 4. Descriptive data

Cluster	Issuer	Price	Return
1	AHAP	62.41667	-0.00894
2	AMAG	306.1667	-0.00394
	ASBI	309.6667	0.009978
	ASDM	222.5	-0.07056
	Average	279.4444	-0.02151
3	ASDM	1052.917	0.004183
	ASMI	882.5	0.070062
	ASRM	882.5	0.070062
	JMAS	852.0833	-0.00013
	Average	917.5	0.036044

Based on Table 4, we can conclude that clusters 1, 2, and 3 are occupied hierarchically by share prices. The issuer has the lowest price in cluster 1, medium price in cluster 2, and high price in cluster 3. Meanwhile, from the aspect of return, most issuers in cluster 3 have positive returns, while clusters 2and one have the majority of negative returns. Referring to (3) which conducts price delay analysis on banking shares against macroeconomic variables, this paper takes data on changes in exchange rates andinterest rates in the year of observation to be regressed using static panel data regression by the least square method. Because panel data regression requires that the number of exogenous variables cannot be greater than the number of cross-sections, only data in clusters 2 and 3 are regressed.

Table 5 and Table 6 show that for exogenous variables significant, clusters 2 and 3 do not have differences. Only the interest rate variable affects the price of insurance shares. In cluster 2, the difference is the interest rate has a positive effect, while in cluster 3, the interest rate is negative. This paper will not discuss further because of the difference in the direction of different interest rates. Still, trials with this regression convince that the k-means clustering method, even though it uses a relatively simple algorithm, is still powerful enough to be used as a generalized analysis tool with a relatively small observation dimension.

Table 5. Fixed (dummy variables) cluster 2

Variable	Coefficient	t-Statistic	Prob.
LOG(ER)	-0.277874	-0.141458	0.8884
LOG(IR)	1.181838	3.129946	0.0038
C	3.558564	5.426100	0.0000

R-squared	0.668977
Adjusted R-squared	0.626264

Table 6. Cross-Section fixed (dummy variables) cluster 3

Variable	Coefficient	t-Statistic	Prob.
LOG(ER)	-0.260282	-0.147309	0.8836
LOG(IR)	-1.455122	-4.284318	0.0001
C	9.331498	15.81859	0.0000
R-squared	0.468191		
Adjusted R-squared	0.404881		

4. Conclusion

This study shows that clustering with *k-means* to determine the characteristics of insurance stocks is quite good. Descriptively, the cluster formed reflects a significant difference in variance, especially forthe price variable. Through the panel data regression test, it is also seen that clusters 2 and 3 have different patterns in responding to changes in interest rates. The regression results reinforce the evidencethat *k-means* clustering can be used to see the initial characteristics of a data set, including inunderstanding the characteristics of issuers.

.

References

[1] Y. Utami, "Indeks Saham Syariah Indonesia: Pergerakan Harga dari Perspektif Asimetri Informasi," *J. Inov. Ekon.*, vol. 4, no. 02, pp. 41–48, 2019, doi: 10.22219/jiko.v4i2.9851.

[2] B. Widagdo and N. R. Satiti, "Indonesian Capital Market Reaction Toward November, 4th 2016 Demonstration in Jakarta," *J. Innov. Bus. Econ.*, vol. 2, no. 01, pp. 29–36, Dec. 2018, doi: 10.22219/JIBE.V2I01.5561.

[3] I. Zuhroh, M. Rofik, and A. Echchabi, "Banking stock price movement and macroeconomic indicators: k-means clustering approach," *http://www.editorialmanager.com/cogentbusiness*, vol. 8, no. 1, 2020, doi: 10.1080/23311975.2021.1980247.

[4] R. Nilavongse, M. Rubaszek, and G. S. Uddin, "Economic policy uncertainty shocks, economic activity, and exchange rate adjustments," *Econ. Lett.*, vol. 186, Jan. 2020, doi: 10.1016/j.econlet.2019.108765.

[5] A. S. Yang and A. Pangastuti, "Stock market efficiency and liquidity: The Indonesia Stock Exchange merger," *Res. Int. Bus. Financ.*, vol. 36, pp. 28–40, Jan. 2016, doi: 10.1016/J.RIBAF.2015.09.002.

[6] K. Lim and C. Hooy, "The delay of stock price adjustment to information: A country-level analysis," *Econ. Bull.*, vol. 30, no. 2, pp. 1609–1616, 2010, Accessed: Dec. 07, 2020. [Online]. Available: https://ideas.repec.org/a/ebl/ecbull/eb-10-00033.html

[7] H. P. Kriegel, E. Schubert, and A. Zimek, "The (black) art of runtime evaluation," *Knowl. Inf. Syst.*, vol. 52, no. 2, pp. 341–378, Aug. 2017, doi: 10.1007/S10115-016-1004-2.

[8] A. Likas, N. Vlassis, and J. J. Verbeek, "The global k-means clustering algorithm," *Pattern Recognit.*, vol. 36, no. 2, pp. 451–461, Feb. 2003, doi: 10.1016/S0031-3203(02)00060-2.

[9] A. K. Jain, "Data clustering: 50 years beyond K-means," *Pattern Recognit. Lett.*, vol. 31, no. 8, pp. 651–666, Jun. 2010, doi: 10.1016/J.PATREC.2009.09.011.