ISSN 2347 - 3983
Volume 8, No. 1 January 2020
International Journal of Emerging Trends in Engineering Research
Available Online at http://www.warse.org/IJETER/static/pdf/file/ijeter09812020.pdf
https://doi.org/10.30534/ijeter/2020/09812020

Application of Data Mining Techniques to Classify World Stock Markets

M Mallikarjuna[1], R Prabhakara Rao[2]
[1,2] Sri Sathya Sai Institute of Higher Learning, Prasanthi Nilayam, AP, India,
[1] mejarimallik@gmail.com [2] rprabhakararao@gmail.com

ABSTRACT

It is imperative for the participants of stock markets to understand the characteristics of stock markets for effective decision making. However, it is difficult to understand the dynamics of every single market, so the classification of stock markets with similar characteristics as a group would be a great help for the stakeholders of the stock markets. There exist some classifications based on various socio-economic and financial factors, which involve huge costs and also this cannot be verified as it is very complex to obtain the data on these variables. There are very few studies in the literature on the classification of markets with quantitative variables. In this context, this study aims at classifying the stock markets into different groups based on their key financial factors. This study considers forty-five stock markets for classification by using data mining techniques, viz. K-Means, Hierarchical, and Fuzzy C-Means. The results show that Fuzzy C-Means clustering is found to be the most suitable method.

Key words: Classification, Data mining, Financial Characteristics, Fuzzy C-Means, Hierarchical, K-Means, Stock markets.

1. INTRODUCTION

After financial liberalization, financial intermediaries have become the crucial institutions in the process of economic growth. Especially, the stock markets have been one of the vital institutions, which have been channeling the surplus funds from the investors to the industries. There are several theoretical as well as empirical studies, which showed that there exists high positive relationship between the economic growth and stock markets [8], [9], [10], [14], [15]. These studies have ascertained that financial development has a significant positive impact on economic growth. Given the significance of stock markets concerning the economic growth and returns on investments, the study of stock markets plays a crucial role in investment decision making.

Moreover, the integration of markets coupled with automated trade facilities stimulated the investors' opportunities to diversify their investments and thereby increasing the rates of return on investments fetching higher profits [5], [6], [23]. With the advancement in information and communication technology, the participants of the stock markets can access real-time data and information about market movements, opinions of experts, and the information of economic, political and social events, which affect the financial markets. Also, with high-speed computing facilities the players of the stock markets have the advantage of making their investment decisions faster and dynamic way.

It is essential for the investors to understand the structure, risk-return characteristics of every stock market for effective investment decision-making [2]. However, understanding every market may not be feasible, as it is a laborious task and a time-consuming process [12]. Nevertheless, one can identify groups of markets with similar features by using appropriate analytical tools. Classification of markets would also be beneficial to the active stakeholders in the stock markets such as institutional and retail investors, fund managers, hedge funds, financial regulators, firms, policymakers and other active players of financial markets. Institutional and retail investors, fund managers, and hedge funds can use this classification to minimize their investment risk and diversify their portfolio and tap the opportunities, which yield higher returns in the world stock markets. This classification can be useful to regulators and policymakers as they can emulate the best management and regulatory practices from the advanced markets to their domestic markets. The firms can approach the new markets apart from domestic markets to raise capital with lesser cost based on their requirements.

There are some international agencies such as International Monetary Fund (IMF), World Bank (WB), and financial service providers like Morgan Stanley Capital International (MSCI), Financial Times Stock Exchange (FTSE), and Dow Jones (DJ), classifying the stock markets based on socio-economic factors, opinion polls, and qualitative variables. However, these qualitative measurements may not reveal full information [18]. According to MSCI [19], if there is misclassification of a market, it is difficult to track the index and it will increase the risk and cost of the misclassification error. Moreover, it is not possible to verify these classifications as it is very expensive and difficult to obtain the data and also these classifications are not being updated dynamically in accordance with the speed at which the changes take place in the investment environment. These are the serious limitations of classification by these agencies.

As an alternative few attempts were made to classify the stock markets by using quantitative financial variables and indices. A study by [3] examined the factors that discriminate the developed and emerging stock markets. The results of their study revealed that the market depth, size of the market, transaction value are more prominent factors that differentiate the financial markets. However, the market activity that comprises of corporate funding, fund capacity, and turnover velocity is not a significant factor for classifying the markets. Recent studies such as [13] and [20] observed that the stylized facts of market returns such as volatility, informational asymmetries play a significant role in differentiating the markets as mature and immature markets. A study by [20] classified the stock markets of 40 countries based on quantitative variables as developed, emerging and frontier markets by using discriminant analysis. According to this study, Israel, Namibia, and Kuwait are misclassified as emerging markets, where Israel is classified as a developed market and Namibia and Kuwait are classified as frontier markets by the agencies. Apart from this, the rest of the classifications are in line with the classification with qualitative variables.

Though these studies classified the stock markets using some quantitative variables and factors, there are some limitations in the methods employed in these studies. For example, [20] used K-Means Clustering, which is sensitive to outliers. Also, a study by [18] employed discriminant analysis to classify the markets, however, this method also has some serious limitations such as this method might not be efficient for the variables, which are distributed non-normally [9].

In recent times, with the advancement in automated collection and storage tools, the availability of data has increased exponentially [1], [25]. On the other hand, the analysis of the huge amount of data demands more sophisticated techniques to draw valid conclusions. This necessitated the invention of new methods to analyze the ever-increasing data. Data mining is one such class of tools, which is used to understand the hidden patterns and analyze the data. Clustering, a technique in data mining has a wide range of applications in finance as well.

From the literature on classification of stock markets, it can be observed that there are very few studies on classification based on quantitative financial characteristics and on the cost-effective classification of stock markets by using the openly available data. In this context, this study has been taken up with an aim to classify the stock markets by considering the quantitative variables such as the size, depth, efficiency, access, and stability of the stock markets. For this classification, we employ the data mining techniques viz. K-Means, Hierarchical, and Fuzzy C-Means clustering methods as these are the most widely used for unsupervised clustering. This study has two distinctive features, one, it is cost-effective as the data has been taken from open sources and two, it is easily verifiable. The rest of the study is divided as: the second section describes the data and methods employed for the study, third section presents the results, followed by the summary and conclusions in the fourth section.

2. DATA AND METHODOLOGY

In this section, we briefly explain the data and variables considered for the study and describe different data mining techniques employed for the classification of the stock markets.

2.1. Data

We have taken five vital factors, which reveal significant information about the stock markets. They are 1. Size: The total market capitalization of the listed companies is taken as a proxy for the market size, which indicates the amount of financial resources in the stock market. Markets with higher market capitalization indicate a bigger size. 2. Depth, which is measured as the ratio of market capitalization to GDP, that compares the total value of all the stocks to the total value of the output of a country. 3. Efficiency, we considered the stock market turnover ratio and liquidity as a proxy for efficiency. Higher the turnover ratio and liquidity, higher the efficiency and lower the turnover ratio and liquidity, lower the efficiency. 4. Stability, which is measured by the volatility in the market. It is an average of the 360-day volatility of the stock index. Lower the volatility, the higher the stability and higher the volatility, the lower the stability of the stock market. 5. Access: The amount of Foreign Portfolio Investments (FPIs) has been taken as a proxy for access, as the amount of foreign investments indicate how open and accessible the markets are. We used the annual data of these variables from the World Bank Database [24] from 2009 to 2016 and calculated the average of all these variables for forty-five countries. The list of countries considered for the study are given in table 1. The variables and their proxies considered for the study are provided in table 2 along with their notations.

Table 1: List of Countries Considered for the Study

1.Australia	24.Mexico
2.Austria	25.Morocco
3.Brazil	26.New Zealand
4.Canada	27.Nigeria
5. Chile	28.Norway
6.China	29.Oman
7.Columbia	30.Peru
8.Cyprus	31.Philippines
9. Egypt	32.Poland
10.Germany	33.Russia
11.Greece	34.Singapore
12.Hong Kong	35.Slovenia
13.Hungary	36. South Africa
14.India	37.South Korea
15.Indonesia	38.Spain
16.Ireland	39.Sri Lanka
17.Israel	40.Switzerland
18. Japan	41.Thailand
19.Jordan	42.Turkey
20.Kazakhstan	43.United Kingdom
21.Luxembourg	44.United States
22.Malta	45.Vietnam
23.Mauritius	

Table 2: Variables and Notations

Factor	Variable	Notation
Size	Stock market capitalization	MCAP
Depth	Stock market capitalization to GDP	DEP
Efficiency	Stock market turnover ratio	TURNR
	Liquidity	LIQ
Stability	Stock price volatility	VOL
Access	Foreign Portfolio Investments	FPI

Source: World Bank database – 2018.

MCAP: Total value of all listed shares in the stock market
DEP: (Stock market capitalization/GDP) *100
TURNR: Total value traded/Stock market capitalization
VOL: The mean value of 360-day volatility in stock prices
LIQ: Total volume of transactions.
FPI: Total Foreign Institutional Investments in the stock markets.

2.2 Data Standardization

Data standardization is one of the vital steps in data pre-processing for any data analysis. In this step, variables or the objects must be scaled, i.e. standardized to make them comparable, before measuring the similarities and dissimilarities among the objects. Especially, when variables are measured on different scales. Generally, variables are scaled to have i) Mean zero and ii) Standard deviation one. The variables (x_i) are standardized using z-score formula, which is given as:

$$z = \frac{x_i - \mu}{\sigma} \quad (1)$$

Where μ is mean and σ is the standard deviation of x_i respectively.

2.3 Determination of Number of Clusters

It is very crucial to choose the optimal number of clusters for a given data set before proceeding to apply clustering algorithms as the results are sensitive to the number of clusters. However, in some circumstances, we can fix the number of clusters, based on the requirement of the study. In this study, we used one of the popular methods, Elbow method [11] for fixing the appropriate number of clusters.

A. Elbow Method

Elbow method considers the total Within Sum of Squares (WSS), which is the sum of the squared deviations from each observation and the cluster centroid, as a function of the number of clusters: The Elbow method involves the following procedure to select the optimal number of clusters:
Step 1. Compute clustering algorithm for different numbers of clusters, i.e. k. For example, 1 to 10 clusters.
Step 2. Calculate total WSS for each k,
Step 3. Plot the curve of WSS for different values of k.
Step 4. The point at which the curve bends is selected as the optimal number for that particular data set.

2.4 Classification Methods

A. Hierarchical Clustering

The hierarchical clustering method is one of the unsupervised clustering algorithms used to group the data based on the similarities among the objects. The algorithm used in the hierarchical method is: Let d_{ij} be the distance between clusters i and j and let cluster i contain n_i objects. Let the set of all remaining d_{ij} be represented as D. Suppose there are N objects to the cluster, the steps involved are:
1. Finding the smallest element d_{ij} that is remaining in D.
2. Merging the clusters i and j into a single new cluster, say k.
3. Calculating a new set of distances, d_{km} using the following distance formula.

$$d_{km} = \alpha_i d_{im} + \alpha_j d_{jm} + \beta d_{ij} + \gamma |d_{im} - d_{jm}| \quad (2)$$

Here m denotes the cluster other than k. The d_{im} and d_{jm} in D are replaced by these new distances. Also, let $n_k = n_i + n_j$
4. Repeating steps 1 to 3, which requires N-1 iterations until D contains a single group made up of all objects. In this method, the clusters are formed in such a way that the pooled WSS is minimized. At each step, two clusters are fused which results in the least increase in the pooled WSS. In this study, we used Ward's Minimum Variance method of agglomerative clustering. The coefficients of the distance equation in Ward Hierarchical Clustering method are:

$$\alpha_i = \frac{n_i + n_m}{n_k + n_m}, \alpha_j = \frac{n_j + n_m}{n_k + n_m}, \beta = \frac{-n_m}{n_k + n_m}, \gamma = 0 \quad (3)$$

B. K-Means Clustering

K-means clustering algorithm is one of the popular unsupervised clustering algorithms because of its effectiveness and less complexity in implementation. K-means clustering algorithm of MacQueen [16] is the most commonly used algorithm. The Euclidian distance measure is used in this algorithm and its objective is to minimize the objective function J, which involves minimizing the distance within the same cluster and maximizing the distance between the clusters which is defined as:

$$J = \sum_{i=1}^{n} \sum_{k=1}^{n} Z_i^k \|x_{ik} - v_i\|^2 \quad (4)$$

Where x_i is the data points, n is the number of observations, v_i is cluster centers, and Z is a membership function, where,

$$Z_i^k = \begin{Bmatrix} 1, & if\ x_i \in cluster\ k \\ 0, otherwise \end{Bmatrix}$$

$$v_i = \frac{\sum_{k=1}^{n} x_{ik} x_{kj}}{\sum_{k=1}^{n} x_{ik} x_{ki}} \quad (5)$$

The steps involved in k-means algorithm steps are:
Step 1. Choosing the number of clusters (k)
Step 2. Selecting the cluster centers, v_i
Step 3. Assigning the data points x_{ik} to the closest cluster
Step 4. Re-computing v_i, cluster centers using equation (5)
Step 5. Repeating steps 3 and 4 until J is invariant (variance $< \varepsilon$)

C. Fuzzy C-Means Clustering

In order to overcome the limitations with K-Means, Bezdek [4] introduced Fuzzy C-means (FCM) clustering method, that is based on Dunn's study [7] which is also considered as an extension of K Means. Fuzzy C Means is a soft clustering method, which allows an object to be associated with different clusters with a varying degree of membership, whereas the hard-clustering methods allow each object to be in only one cluster. The clusters are created based on the distance among the data points, with each cluster having a center. In fact, in FCM, n clusters are formed from a data set with each point in a data set is related to every other cluster as well with a certain degree of association with that particular cluster. A data point with a high degree of association lies closer to the center of the cluster and the data point with a low degree of association lies distant from the center of the cluster.

Fuzzy C-Means clustering involves the following steps:

Initially, one has to fix c, where c is $(2 \leq c < n)$ and then a value for the parameter 'm' must be selected, followed by initialization of the partition matrix $U^{(0)}$. In this algorithm, every step would be specified as 'r', where r= 0, 1, 2.

Step 1. Calculating the c center vector $\{V_{ii}\}$ for each step

$$v_{ij} = \frac{\sum_{k=1}^{n} (\mu_{ik})^m x_{kj}}{\sum_{k=1}^{n} (\mu_{ik})^m} \qquad (6)$$

Step 2. Calculate the distance matrix $D_{[c,n]}$

$$D_{ij} = \left(\sum_{j=1}^{m} \left(x_{kj} - v_{ij} \right)^2 \right)^{1/2} \qquad (7)$$

Step 3. Updating the partition matrix for the rth step, $U^{(R)}$) as

$$\mu_{ij}^{r-1} = \left(1 \Big/ \sum_{j=1}^{c} \left(d_{ik}^{r} / d_{jk}^{r} \right)^{2/m-1} \right) \qquad (8)$$

If $\|U^{(k+1)} - U^{k}\| < \delta,$ then one has to stop, or else one must return to step 2 by updating the membership grades for data points and cluster centers iteratively. In FCM, the cluster centers are moved iteratively to the precise location within a dataset. The underlying phenomena for the FCM clustering techniques is that of the fuzzy behavior, which provides a natural approach to create clusters, with membership weights that have a natural interpretation rather than probabilistic interpretation.

2.5 Validation of Clustering Techniques

It is very important to measure the accuracy of a clustering technique to understand its performance [26]. There exist some methods such as the Dunn index, Rand index, and Silhouette width to validate the clustering techniques. However, several studies used the Silhouette width as it is one of the most widely used reliable methods as well as easily interpretable [11], [17], [21].

A. Silhouette Width

It is very essential to test the validity of the clustering results. One of the commonly used statistics to validate the clustering results is the Silhouette width [22] which measures how accurately an observation is clustered and estimates the average distance between clusters. The silhouette width S_i is calculated for each observation i, by using the formula:

$$S_i = \frac{1}{n} \sum_{i=1}^{n} \frac{b_i - a_i}{\max\,(a_i, b_i)} \qquad (9)$$

Where, a_i is the mean distance between i and all other points within a cluster, n is the total number of points, and b_i is the minimum of the average distances between i and the points in other clusters.

The steps to calculate the Silhouette index are as follows:

Step 1. Calculate the average dissimilarity a_i between each data point i and all other data points of the cluster, in which i is a member.

Step 2. Calculate the average distance d(i,C) of i for all other clusters C, in which, i is not a member. Define the smallest of these dissimilarities as b_i = min C d(i,C), which is the dissimilarity between i and its neighboring cluster, to which it does not belong.

Step 3. In the last step, the silhouette width of i can be obtained by using equation 9. After obtaining the Silhouette width, it is easy to interpret. The observations with a high S_i i.e. near to 1, indicate that they are clustered very well. A small S_i, near to 0, denotes that, the data point lies between two clusters, and the observations that are wrongly clustered will have a negative S_i.Then, we can get the average Silhouette width for a particular clustering method, which can be used to validate the method and compare it with other methods. The range of average Silhouette width is -1 to +1. A high value, i.e. close to 1 indicates that the objects are perfectly clustered. A low value, near to 0 reveals that the objects are poorly clustered. A negative value shows that the objects are placed in the wrong cluster.

3. EMPIRICAL RESULTS

In this section, we present the empirical results, i.e. clustering outputs of three methods employed in the study. We used the Elbow method to determine the optimal number of clusters. The output of the Elbow method is given in figure 1, from which we can observe that the optimal number of clusters is 'three'. Based on these results, we fixed the value of *k* as 3 in clustering methods. Having determined the ideal number of clusters from Elbow method, we applied three widely used clustering methods, viz, Agglomerative or Hierarchical, K-Means, and Fuzzy C- Means methods to classify the stock markets with similar features and we tested the goodness-of-fit of these methods by using the validation technique, Silhouette width. Here, we obtained the average Silhouette width of 0.8311 for Fuzzy C-Means, 0.6111 for Hierarchical and 0.3823 for K-Means methods. Hence, the Fuzzy C-Means method is the most suitable method among the three methods considered in the study. After validating the clustering methods, we plotted the cluster plots to visualize the clusters. The cluster plots from the three methods can be observed in figure 2 to figure 4.

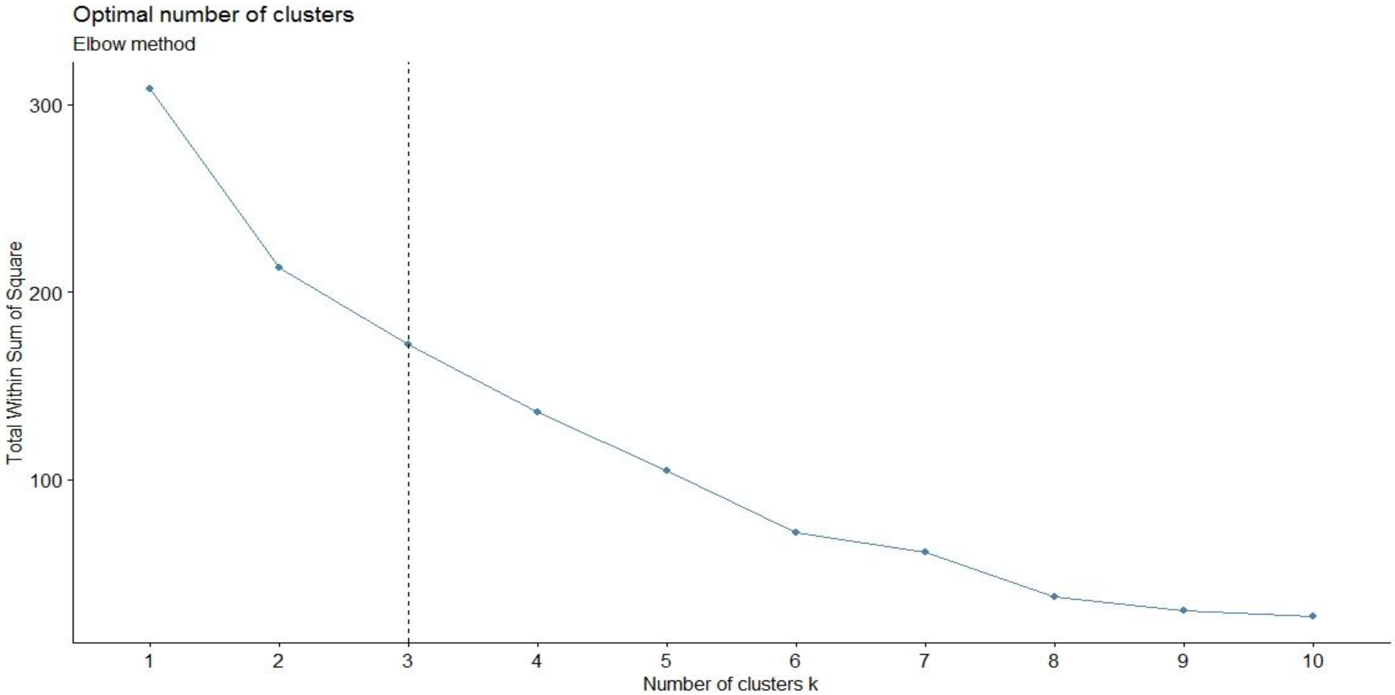

Figure 1:Optimal Number of Clusters from Elbow Method

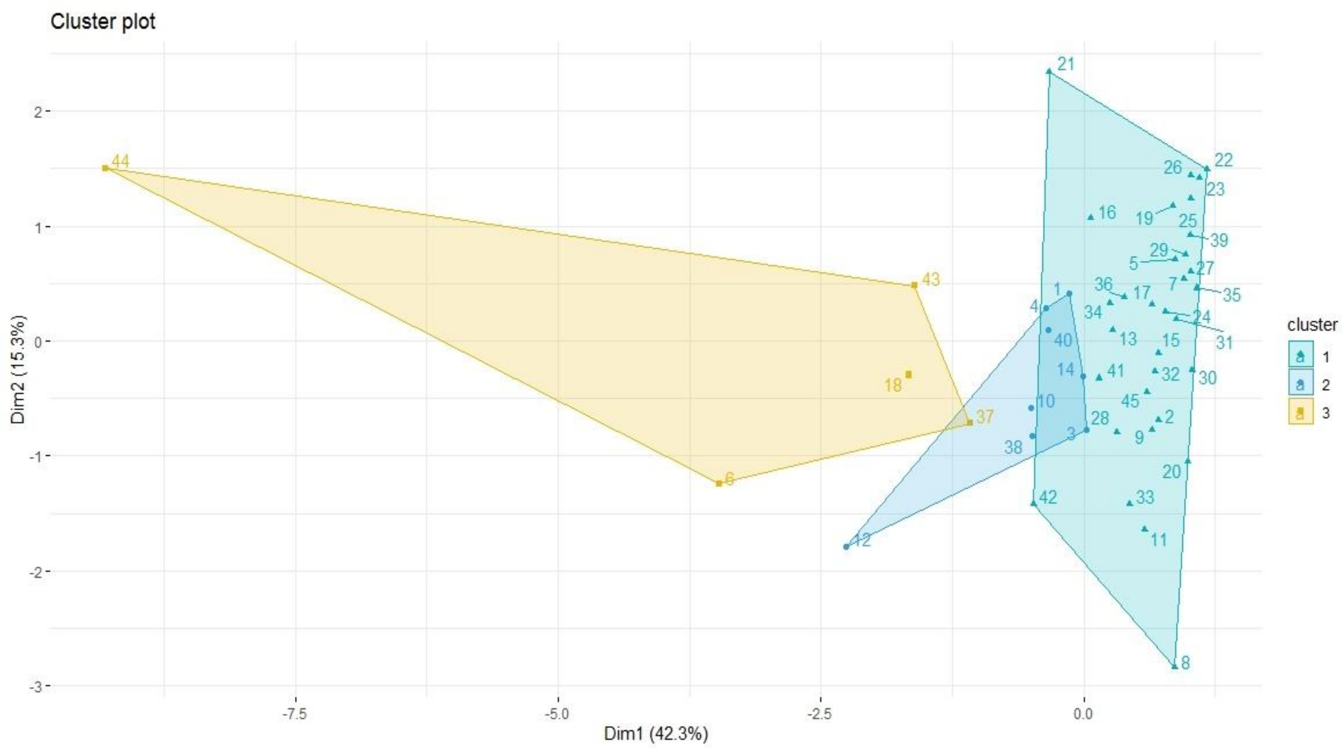

Figure 2: Cluster plot with Hierarchical Method

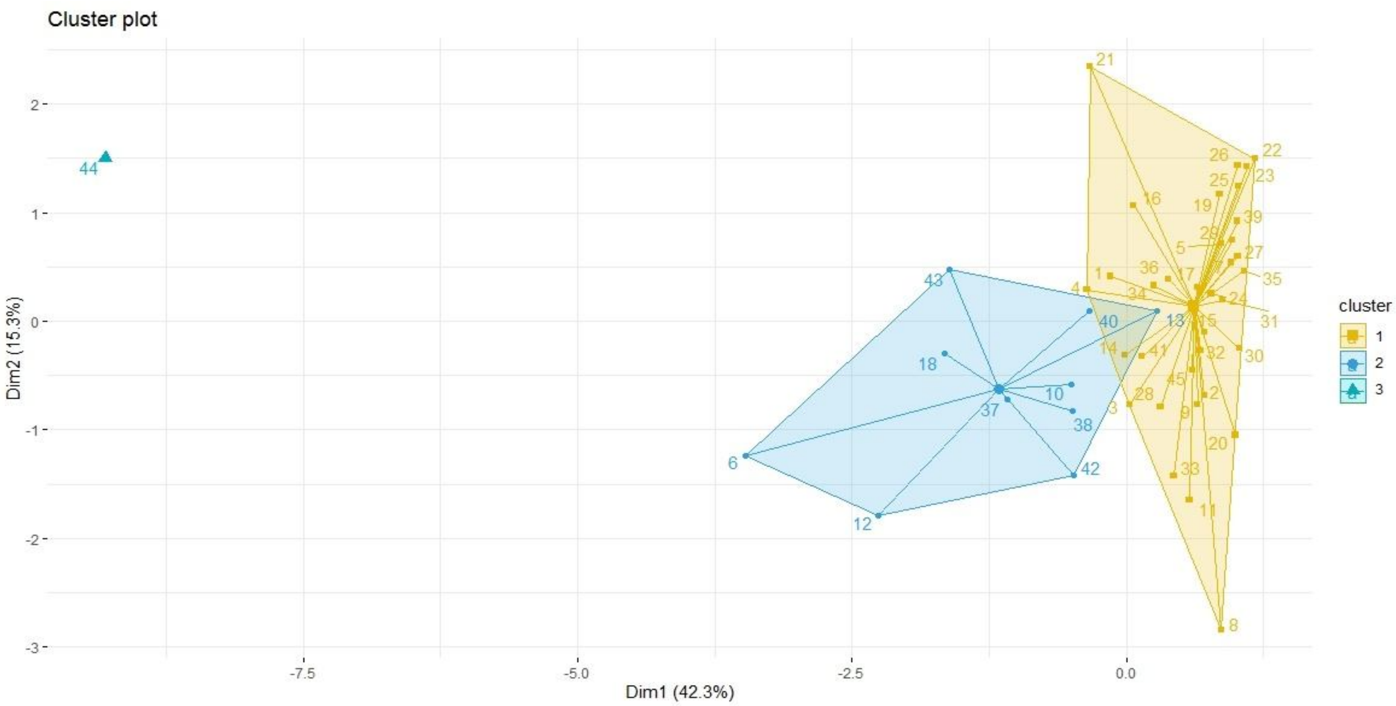

Figure 3: Cluster plot with K-Means Method

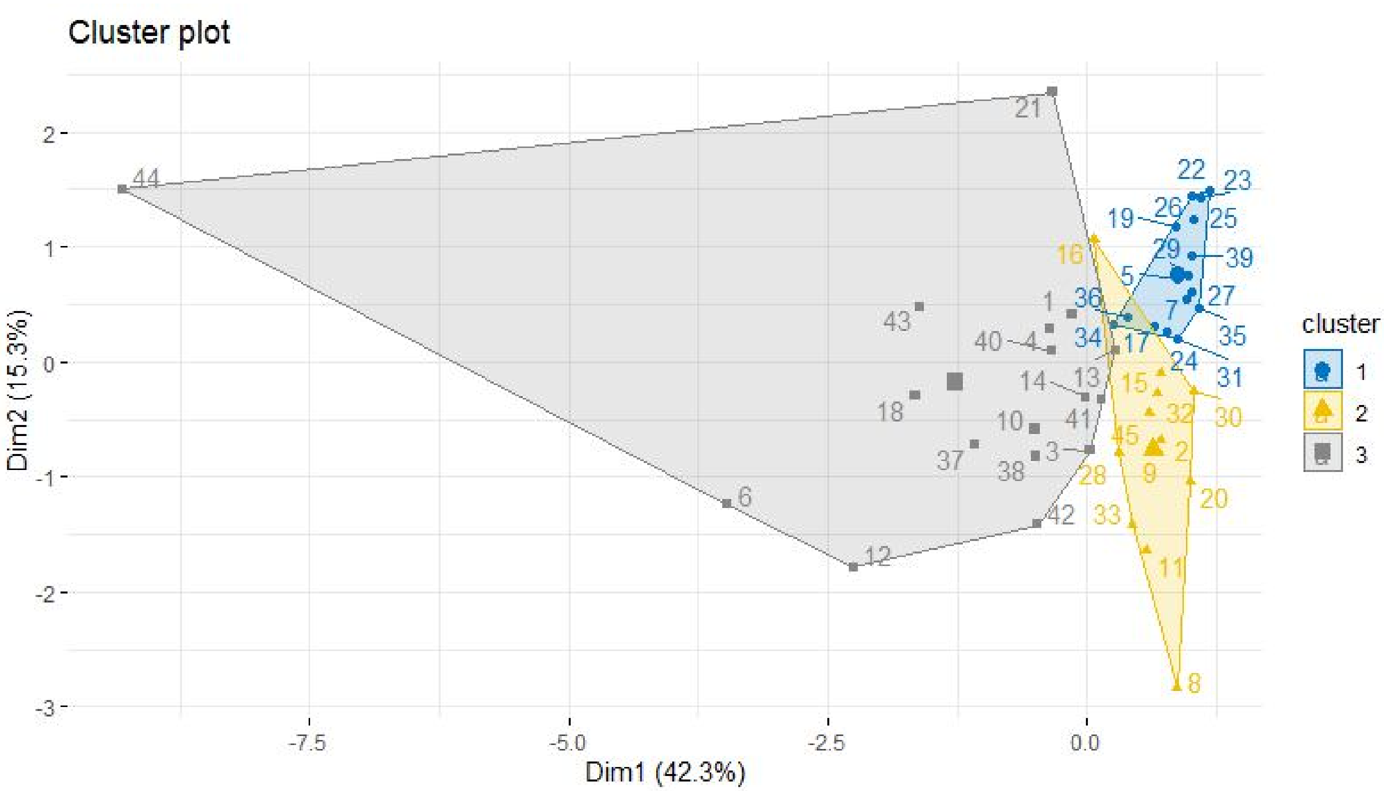

Figure 4: Cluster plot with Fuzzy C- Means Method

From the Fuzzy C-Means cluster plot in fig. 4, there are thirteen markets in cluster 1, namely, Chile, Colombia, Israel, Jordan, Malta, Mexico, Mauritius, Morocco, Nigeria, New Zealand, Oman, Philippines, and Sri Lanka. Cluster 2 consists of fifteen markets, viz. Austria, Cyprus, Egypt, Greece, Indonesia, Ireland, Kazakhstan, Norway, Poland, Peru, Russia, Slovenia, Singapore, South Africa, and Vietnam. In cluster 3, there are seventeen markets, namely, Australia, Brazil, China, Canada, Germany, Hong Kong, Hungary, India, Japan, Luxembourg, South Korea, Spain, Switzerland, Thailand, Turkey, UK, and the US. One interesting observation from the results of all the three clustering methods is that the distance between the US stock markets and all other markets is very high, indicating that the stock markets of US are way ahead of all other markets in terms of size, efficiency, depth, liquidity, and stability. The list of countries in each of the clusters is given in table 3.

Table 3: Clusters of Stock Markets

Cluster 1(13)	**Cluster 2 (15)**	**Cluster 3(17)**
5. Chile	2. Austria	1. Australia
7. Colombia	8. Cyprus	3. Brazil
17. Israel	9. Egypt	4. Canada
19. Jordan	11. Greece	6. China
22. Malta	15. Indonesia	10. Germany
23. Mauritius	16. Ireland	12. Hong Kong
24. Mexico	20. Kazakhstan	13. Hungary
25. Morocco	28. Norway	14. India
26. New Zealand	30. Peru	18. Japan
27. Nigeria	32. Poland	21. Luxembourg
29. Oman	33. Russia	37. South Korea
31. Philippines	34. Singapore	38. Spain
39. Sri Lanka	35. Slovenia	40. Switzerland
	36. South Africa	41. Thailand
	45. Vietnam	42. Turkey
		43. UK
		44. US

4. SUMMARY AND CONCLUSIONS

One of the crucial prerequisites for a better investment decision-making process is having the knowledge of characteristics of stock markets. As it is difficult to comprehend each and every market in the construction of the portfolio for investors, classification of stock markets with similar characteristics would be useful for market participants. There exist few classifications by some agencies such as Dow Jones (DJ), Financial Times Stock Exchange (FTSE), and Morgan Stanley Capital International (MSCI), by using financial characteristics. However, these classifications involve huge cost and also, they are not verifiable. This study makes an attempt to classify the stock markets based on their quantitative financial features such as size, access, depth, efficiency, and stability by employing data mining techniques.

In this study, we considered the average of annual data of the chosen variables for the period 2009 to 2016 for forty-five stock markets. We employed three unsupervised clustering methods viz., Hierarchical, K-Means, and Fuzzy C-Means clustering for classifying the stock markets. First, we selected the optimum number of clusters as three, as determined by employing the Elbow method. After forming the clusters, we used the average Silhouette width to validate and compare clustering methods. The results of this study suggest that the Fuzzy C- Means clustering method is the most appropriate for the classification of stock markets. These results show that there are thirteen markets in cluster 1, namely, Chile, Colombia, Israel, Jordan, Malta, Mexico, Mauritius, Morocco, Nigeria, New Zealand, Oman, Philippines, and Sri Lanka. There are fifteen markets in cluster 2, viz. Austria, Cyprus, Egypt, Greece, Indonesia, Ireland, Kazakhstan, Norway, Poland, Peru, Russia, Slovenia, Singapore, South Africa, and Vietnam. Cluster 3 consists of seventeen markets, namely, Australia, Brazil, China, Canada, Germany, Hong Kong, Hungary, India, Japan, Luxembourg, South Korea, Spain, Switzerland, Thailand, Turkey, UK, and the US.

REFERENCES

1. A. Dhankhar, and K. Solanki. **A comprehensive review of tools and techniques for big data analysis.** *International Journal of Emerging Trends in Engineering Research*, Vol. 7, No. 11, pp. 556-562, 2019. https://doi.org/10.30534/ijeter/2019/257112019
2. P. Alagidede. **Return behavior of Africa's emerging markets**, *The Quarterly Review of Economics and Finance*, Vol. 51, pp. 133-140, 2011. https://doi.org/10.1016/j.qref.2011.01.004
3. G. Alrgibi, M. Ariff, and L. Murray. **What factors discriminate developed and emerging capital markets?** *Applied economic letters*, Vol. 17, No. 13, pp. 1293-1298, 2010. http://dx.doi.org/10.1080/00036840902881850
4. J. C. Bezdek. **Pattern recognition with fuzzy objective function algorithms**, New York, Plenum Press, 1981. http://dx.doi.org/10.1007/978-1-4757-0450-1
5. R. F. Bruner, W. Li, M. Kritzman, S. Myrgren, and S. Page. **Market integration in developed and emerging markets: Evidence from the CAPM**. *Emerging Markets Review*, Vol. 9, pp. 89-103, 2008. http://dx.doi.org/10.1016/j.ememar.2008.02.002
6. S. T. Cavusgil, P. N. Ghauri, and M. R. Agarwal. **Doing business in emerging markets**, 2nd ed, Sage Publications, 2012. http://dx.doi.org/10.4135/9781483328720

7. J. C. Dunn. **A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters**. *Journal of Cybernetics*, Vol. 3, No. 3, pp. 32-57. 1973. http://dx.doi.org/10.1080/01969727308546046
8. R. Fisman, and I. Love. **Trade credit, financial intermediary development and industry growth**. *Journal of Finance*, Vol. 58, pp. 353-374. 2003. https://doi.org/10.1111/1540-6261.00527
9. C. Fraley and A. E. Raftery. **Model-based clustering, discriminant analysis, and density estimation**, *Journal of the American Statistical Association*, Vol. 97, No. 458, pp. 611 – 631, 2002. https://doi.org/10.1198/016214502760047131
10. S. K. Guptha, and R. P. Rao. **The causal relationship between financial development and economic growth: an experience with BRICS economies**. *Journal of Social and Economic Development*, Vol. 20, No. 2, pp. 308–326, 2018. http://dx.doi.org/10.1007/s40847-018-0071-5
11. A. Kassambara. **Practical guide to cluster analysis in R,** 1st ed, Statistical Tools for High-Throughput Data Analysis (STHDA), Create Space Independent Publishing Platform, 2017. Available at: https://www.datanovia.com/en/product/practical-guide-to-cluster-analysis-in-r/?url=/5-bookadvisor/17-practical-guide-to-cluster-analysis-in-r/
12. G. Kohers, N. Kohers, and T. Kohers. **The risk and return characteristics of developed and emerging stock markets: the recent evidence.** *Applied Economics Letters*, Vol. 13, No. 11, pp. 737-743, 2006. https://doi.org/10.1080/13504850500407210
13. H. Krichene. **Immature versus mature stock markets' properties: univariate and multivariate stylized facts analysis**. Hyogo, Japan, 2016.
14. R. Levine, and S. Zervos. **Stock markets, banks, and economic growth**. *American Economic Review*, Vol. 88, No. 3, pp. 537–558. 1998. Available at: https://www.jstor.org/stable/116848?seq=1
15. B. McCaig, and T. Stengos. **Financial intermediation and growth: some robustness results.** *Economic Letters*, Vol. 88, No. 3, pp. 306–312. 2005. https://doi.org/10.1016/j.econlet.2004.12.031
16. J. B. MacQueen. **Some methods for classification and analysis of multivariate observations**. *Proceedings of 5-th Berkeley Symposium on Mathematical Statistics and Probability*, Berkeley, pp. 281-297, 1967. Available at: https://projecteuclid.org/euclid.bsmsp/1200512992
17. G. O. Meara. **Mining and classifying images from an advertisement image remover**. *Annals of Data Science*, Vol. 6, No. 2, pp. 279-303. 2019. https://doi.org/10.1007/s40745-018-0164-1
18. B.V. Mendes, and R.A. Martins. **Determinants of stock market classifications**. *Applied Economics Letters*, Vol. 25, No. 17, pp. 1244-1249, 2018. https://doi.org/10.1080/13504851.2017.1414927
19. MSCI. MSCI Announces the Results of Its Annual Market Classification Review. Available at: https://www.msci.com/market-classification. 2018.
20. P.S. Neto, G.D. Cavalcanti, F. Madeiro, and T. A. Ferreira. **An ideal gas approach to classify countries using financial indices.** *Physica A*, Vol. 392, pp. 177-183, 2013. https://doi.org/10.1016/j.physa.2012.07.049
21. U. Rani, and S. Sahu. **Comparison of clustering techniques for measuring similarity in articles**. *IEEE International Conference on Computational Intelligence and Communication Technology*, pp. 1-7, 2017.
22. P. J. Rousseeuw. Silhouettes: **A graphical aid to the interpretation and validation of cluster analysis**. *Computational and Applied Mathematics*, Vol. 20, pp. 53-65., 1987. https://doi.org/10.1016/0377-0427(87)90125-7
23. A. Singh. **Stock markets in low and middle-income countries**. Cambridge: Centre for Business Research, University of Cambridge, 2008. Available at: https://pdfs.semanticscholar.org/1ff1/c06d95fbe81df635dd60f26145732bdb3108.pdf
24. World Bank. **World Bank Open Data**. Available at: https://data.worldbank.org/
25. P. Hooda, and P. Mittal. **An exposition of data mining techniques for customer churn in telecom sector**. *International Journal of Emerging Trends in Engineering Research*, Vol. 7, No. 11, pp. 506-511, 2019. https://doi.org/10.30534/ijeter/2019/177112019
26. A. T. Teologo, and L. Materum. **Cluster-wise Jaccard accuracy of KPower Means on multipath datasets**. *International Journal of Emerging Trends in Engineering Research*, Vol. 7, No. 8, pp. 203-208, 2019. https://doi.org/10.30534/ijeter/2019/16782019