# Clustering Algorithms for Risk Management

Ivana P. Marković[1], Jovica M. Stanković[2], Jelena Z. Stanković[3], Milos B. Stojanović[4]
Faculty of Economics at the University of Ni, Trg kralja Aleksandra Ujedinitelja 11
18000 Ni, Serbia
{ivana.markovic@eknfak.ni.ac.rs} {jovica.stankovic@eknfak.ni.ac.rs} {jelenas@eknfak.ni.ac.rs}
{milos.stojanovic@vtsnis.edu.rs }

**ABSTRACT:** *The investments in capital market depends on deeper analysis of yield and risk. The investments depend on the stock market which improve the interest of the people. We intend to study the stock market using cluster algorithms. We found that the cluster algorithms can enable to identify groups of stocks in the existing portfolios and help to the risk diversification.*

## 1. Introduction

Portfolio optimization represents one of the most important aspects of making investment decisions. Despite the fact that it was presented to the academic and professional audiences back in the 1950s, modern portfolio theory is the first quantitative framework for the diversification of investment portfolios, which is widely being used today. The simplified initial assumptions in terms of investor preferences and the probability distribution of the financial assets' returns have made it possible to form a model of optimization, which, according to Markowitz's approach, is based only on the values of the expected return and risk. A rational investor will prefer efficient portfolios, which provide maximal return for an acceptable level of risk. However, in cases when a large number of financial instruments are being considered, high dimensionality can prevent the precise evaluation of a complex correlation structure and risk. Thus, the application of Markowitz's model in the optimization of large portfolios can result in the allocation of financial means to suboptimal investment alternatives.

The problem of diversification when creating effective portfolios in comparison to more conventional methods could also be solved by the application of numerous alternative methods. In numerous studies, the algorithms of machine the most frequently used approaches include artificial neural networks [1], [2], genetic algorithms [3], and cluster analysis [4] and [5].

Cluster analysis represents a technique of supervised machine learning which attempts to establish previously unknown and useful connections among the data, whereby different insight into the structure of the data is obtained. A special type of clustering is the clustering of time series, which are classified as dynamic data since their characteristics can change over observed period of time. Additionally, clustering of financial time series is especially challenging task, taking into consideration the fact that the financial market is a complex, evolving and dynamic system whose behavior is pronouncedly non-linear.

Most of the research that focuses on cluster analyses with the aim of portfolio optimization refers to the developed financial markets, as shown in [5] and [6]. On the other hand, the presented empirical study focuses on developing markets, and the Belgrade Stock Exchange in particular. This aim of this paper is to apply a cluster algorithm in the identification process of various groups of shares, whose inclusion in the portfolio would provide risk diversification. The subject matter of the analysis is portfolio optimization of the shares which are constituents of the BELEX15 and BELEX*line* index.

The paper is organized as follows: the second part presents the theoretical basis of the applied clustering methods. The third describes the used data and the experimental framework, while the fourth part presents the results of the cluster analysis and the results of the portfolio optimization. In the final part, some of the conclusions and directions for further research are presented.

## 2. Clustering Algorithms

This paper analyzes the influence of two most frequently used groups of clustering algorithms, partitional and hierarchical clustering algorithms, on portfolio optimization. In partitional clustering, the dataset object is divided into non-overlapping clusters, while in the case of hierarchical clustering, nested clusters are organized as subclusters.

### 2.1 Partitional Clustering Algorithms
K-means is one of the most widely used algorithms in this group thanks to its easy implementation, simplicity, efficiency, and empirical success [7].

Let $X = \{x_i\}$ 1,..., $n$ be the set of $n$ $d$-dimensional points to be clustered into a set of $K$ clusters, $C = \{Ck, k = 1, ... , K\}$. Let $\mu_k$ be the mean of cluster $C_k$ while $x_i$ is a data point belonging to the cluster $C_k$. The squared error between $C_k$ and the points in cluster $C_k$ are defined as:

$$W(C_k) = \sum_{x_i \in C_k}(x_i - \mu_k)^2 \tag{1}$$

Each observation ($x_i$) is assigned to a given cluster such that the sum of squares (*SS*) distance of the observation to their assigned cluster centers $\mu_k$ is minimal. The total within-cluster variation is defined as [8]:

$$tot.withinss = \sum_{k=1}^{K} W(C_k) = \sum_{k=1}^{K}\sum_{x_i \in C_k}(x_i - \mu_k)^2 \tag{2}$$

The *total within-cluster sum of square* measures the compactness of the clustering, and it should be minimized.

A *K*-means algorithm starts with an initial partition with K clusters, by randomly selecting $k$ objects from the data set as the initial cluster centers or means, and assigns observations to the closest cluster centers. After all the observations are assigned to the clusters, new cluster centers are computed, and the distance between each observation and new obtained cluster centers is recalculated. These steps are repeated until the cluster assignments stop changing or the maximum number of iterations is reached so that the total within-cluster variation between points in the cluster and the corresponding centroid is minimized.

### 2.2 Hierarchical Clustering
In hierarchical clustering, the clusters are determined by applying an agglomerative or divisive algorithm. Agglomerative clustering uses a bottom-up approach to perform hierarchical clustering, so that each observation is primarily found in its special cluster, and then the clusters are combined into larger clusters until finally all the observations belong to the same

cluster or until a stopping criterion is satisfied. A divisive algorithm generates clusters following the opposite procedure, that is in a top-down manner, beginning with one cluster which contains all the observations and then later determining the subclusters.

In this paper we used two categories of hierarchical clustering, the principal agglomerative clustering algorithms single linkage and complete linkage, which use linkage criteria to determine the metric which is used when grouping clusters. The single linkage algorithm calculated the minimum distances between all observations of pairs of clusters, while the complete linkage algorithm calculated the maximum distance between observations of pairs of clusters.

A similarity measure is of essential importance for each clustering algorithm, but similarity measures on time series data are much harder to determine than on constant data, because of their own continuous nature. This is why the similarity measure is, as a rule, carried out in an approximate manner [9]. When clustering time series, various types similarity measures are used, considering that each measure reflects a different type of similarity among time series data.

The shape-based measures as the Euclidean distance and Dynamic Time Warping (DTW) [10], which find similarities in the time series in the domain of time and shape, are widely used for clustering of times series data [9]. This is why in the experimental part of this study we used these two specific measures.

The Euclidean metric is defined as the root of the square difference between co-ordinates of pairs of objects [8] and the Euclidean distance between two points $X_i$ and $X_j$, with the $m$ dimension calculated as:

$$\text{Dist} X_i X_j \ = \ \sqrt{\sum_{k=1}^{m} (X_{ik} - X_{jk})^2} \qquad (2)$$

On the other hand, DTW aligns two time series., $Q = q_1, q_2,...,q_n$ and $P = p_1, p_2,...,p_m$, using matrix $M_{nxm}$, in order to minimize their difference. The element $(i, j)$ of the matrix $/$ represents the distance $d = (q_i, p_j)$ between two points $q_i$ and $q_j$ where in order to calculate the distance various methods can be used. In this paper we used a standard Euclidean distance. The warping path, denoted by $W$, $W = w_1, w_2,...w_k...w_K$ while $max(m,n) \le K \le m + n - 1$, actually represents a group of matrix elements which define the mapping between $Q$ and $P$, with the $k$-th element $w_k = (i_k, j_k)$. The wrapping path needs to satisfy three conditions: the continuity condition, the boundary condition and the monotonicity condition [10]. The optimal path is determined by applying dynamic programming in order to find the wrapping path $W$ which minimizes the wrapping cost, that is, minimize the distance between the two time series:

$$DTW(Q,P) = \min_{W} [\sum_{k=1}^{K} d(w_k)] \qquad (3)$$

Where $d(w_k) = (q_{i_k}, p_{i_k}) = (q_{i_k}, p_{i_k})^2$. The complete derivations can be found in [11].

## 3. Experimental Data

The value of BELEX*line* and BELEX15 indexes are determined by the prices of the most liquid shares which are continuously being traded on the regulated market of the Belgrade Stock Exchange. An adequate evaluation of the basic characteristics of financial time series requires a certain duration of the financial time series, which in the opinion of analysts reduces the number of available stocks to 29. Taking into consideration the required conditions which the companies issuing the shares need to meet in order for them to be included in these lists, the starting point in this study was the assumption that they are the most liquid securities on the Belgrade Stock Exchange.

The data used in this paper were taken from the website of the Belgrade Stock Exchange (www.belex.rs). The series of values of the index, as well as of individual shares, include records from January 1, 2008 to December 31, 2017, which in total includes 2524 days of training. Clustering algorithms are executed using 29 assets observation each of which with the 2524 trading data.

Afterwards, we distributed assets inside each cluster relying on the Omega ratio as the optimization strategy for asset allocation. The Omega ratio allows comparisons between risk and return at different threshold levels for various asset choices

[12]. Assuming that $F(x)$ is a cumulative distribution function of the rate of return on investment and $r_m$ is the minimum acceptable return, the Omega ratio can be calculated using the following equation:

$$\Omega_t = \frac{\int_{r_m}^{\infty}(1-F(x))dx}{\int_{-\infty}^{r} F(x)dx} \qquad (5)$$

As a general measure for the evaluation of the quality of the obtained cluster analysis, in this paper we used three different economic performance measures: the return $r_t$, risk $\sigma_t$ and the Sharpe ratio $SR_t$, which are calculated according to the following equations:

$$r_t = logCP_t - logCP_{t-1} \qquad (6)$$

$$\sigma_t = \sqrt{\frac{\sum_{t=1}^{n}(r_t - \bar{r})^2}{n-1}} \qquad (7)$$

$$SR_t = \frac{r_t}{\sigma_t} \qquad (8)$$

where $CP$ represents closing price at $t$, $t = 1, 2, ..., n$

## 4. Results and Discussion

In this paper we tested the potential of clustering in portfolio optimization using two different clustering algorithms: K-means with an Euclidean distance, as a representative of partitional clustering, and a single linkage (SL) and a complete linkage (CL) agglomerative clustering algorithm by applying the DTW metric.

Figure 1 shows the dendrogram of hierarchical clustering by using a single-linkage method on the data set, while Figure 2 shows the dendrogram for the same time period but with the application of a complete linkage method.
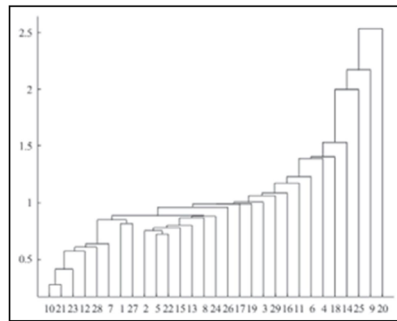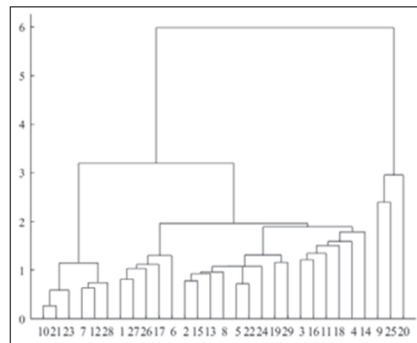


Figure 1. Single linkage



Figure 2. Complete linkage

According to the figures, we can note a difference in the content and distribution of the clusters, depending on the applied methods.

Figure 3. shows the results of the application of the DTW metric on the studied group for two shares AGBC and VBAV over a three-month period.
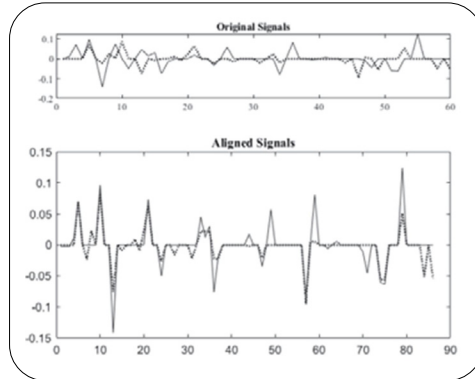


Figure 3. Example of DTW alignment between AGBC (1) and VBAV(27) shares in a three-month period

It is known that the performance of each clustering algorithm is affected by the number of created clusters. In this paper we selected the optimal number of clusters using the Calinski-Harabaz index [13]. This score is given as ratio of means between clusters and the within-cluster dispersion. Typically, for K-means the number of clusters is determined using the Elbow method by computing K-means clustering using different values of K. Next, the wss (within sum of square) is plotted according to the number of clusters. The location of a bend (knee) in the plot is generally considered as an indicator of the appropriate number of clusters.

| Clustering Algorithms | Shares |
|---|---|
| K-means ($K = 4$) | JMBN, GLOS, TIGR, PRGS |
| CL ($K = 2$) | PRGS, JMBN |
| SL ($K = 2$) | PRGS, GLOS |
| BELEX15 | NIIS, KMBN, AERO, FITO, MTLC, TGAS, ALFA, ENHL, JESV, SJPT, IMPL |
| BELEX*line* | AERO, NIIS, MTLC, KMBN, FITO, JMBN, TGAS, ALFA, ENHL, FINT, EPEN, JESV, LSTA, SJPT, IMPL, DNOS, KOPB, EPIN, TIGR, VDAV, GLOS, VBSE |

Table 1. The selection of shares based on the algorithm and model

Based on the data shown in Table 1, we can note that the *K*-means algorithm a greater number of clusters, 4 compared to 2 or 3 which is the number of clusters obtained by applying the method of hierarchical clustering. In Table 1 we can also note a difference in the selected shares based on the applied algorithms and models of clustering. The table presents the shares of both market indexes, BELEX15 and BELEX*line*. In relation to the market indexes, we can note that the clustering algorithms recommend the use of a smaller number of shares. The results obtained by the portfolio optimization based on applying the clustering method, compared to the values of the market indexes BELEX15 and BELEX*line* during the monitored test period are represented in tables 2 and 3.

Based on the first test period, which includes the first quartal of 2018, we can note that the K-means algorithm allow the creation of a portfolio which is characterized by a significantly smaller risk, but a less favorable relationship between return and risk, measured by the Sharpe and Omega ratios, compared to the portfolios obtained by using algorithms of hierarchical clustering. The complete linkage method is used to create a portfolio which realizes the best results based on all the

|  | Return | Risk | Sharpe | Omega |
|---|---|---|---|---|
| K-means | -0.214 | 0.214 | -0.998 | 0.696 |
| CL | -0.191 | 0.273 | -0.699 | 0.767 |
| SL | -0.233 | 0.244 | -0.955 | 0.705 |
| BELEX15 | -0.299 | 0.177 | -1.689 | 0.525 |
| **BELEX***line* | **-0.222** | **0.100** | **-2.226** | **0.428** |

Table 2. Test period 01.01.-01.03.2018

suggested algorithms, but also market indexes, considering that the Sharpe ratio of such a portfolio is 2.5 times smaller than the BELEX15 index, that is, 3 times smaller than the BELEX*line* index.

|  | Return | Risk | Sharpe | Omega |
|---|---|---|---|---|
| K-means | -1.309 | 0.401 | -3.265 | 0.551 |
| CL | -1.042 | 0.504 | -2.066 | 0.676 |
| SL | -1.417 | 0.567 | -2.500 | 0.609 |
| BELEX15 | -1.460 | 0.396 | -3.687 | 0.479 |
| **BELEX***line* | **-1.193** | **0.257** | **-4.648** | **0.387** |

Table 3. Test period 01.01.-31.12.2018

During the second test period, which includes all of 2018, we can note that all the clustering models provide better results for the studied evaluation parameters compared to the values of the market indexes BELEX15 and BELEX*line*. The application of the K-means algorithm enables the creation of portfolios whose performances are improved compared to those of the studied market indexes in terms of the return and risk ratio, while the complete linkage method provides a portfolio which offers the most favorable ratio between return and risk and significantly improves the features of the optimized portfolio.

Experiments have shown that the traditional Euclidean distance metric cannot always be a suitable distance function in financial time series analysis, while DTW is able to find optimal global alignment between sequences and capture specific similarities.

## 5. Conclusion

Considering that learning from data and adapting to financial market changes are a specific and favorable advantage of machine learning methods, we have seen an increased use of these methods in financial analyses. As a special type of machine learning, clustering algorithms have also recently become the focus of practitioners, as these algorithms can both identify and incorporate the fast changes of dynamic financial markets.

The results obtained in this study indicate the possibility of creation a portfolio by using hierarchical clustering algorithms in the portfolio selection task, which would result in improved performances compared to the portfolios of the market indexes in the sense of a stable diversified portfolio that shows a lower risk measured in terms of return volatility.

In the future, we plan to investigate the effect of expanding the group of shares used to create a portfolio on the performance of a clustering algorithm, testing the adequate timing for rebalancing a portfolio, analyzing other strategies for the selection of shares from the cluster which are included in the realization of the portfolio and analyzing the use of more advanced

algorithms for clustering which might additionally improve the clustering of financial time series with the aim of portfolio optimization.

## References

[1] Fernández, S., Gómez, (2007). Portfolio selection using neural networks, *Computers & Operations Research*, 34 (4) 1177-1191. Available: 10.1016/j.cor.2005.06.017

[2] Ko, P., Lin, P. (2008). Resource allocation neural network in portfolio selection, *Expert Systems with Applications*, 35 (1-2) 330-337.

[3] Chang, T., Yang, S., Chang, K. (2009). Portfolio optimization problems in different risk measures using genetic algorithm, *Expert Systems with Applications*, 36 (7) 10529- 10537. Available: 10.1016/j.eswa.2009.02.062.

[4] Nanda, S. R., Mahanty, B., Tiwari, M. K. (2010). Clustering Indian stock market data for portfolio management, *Expert Systems with Applications*, 37 (12) 8793-8798.

[5] Tola, V., Lillo, F., Gallegati, M., Mantegna, R. (2008). Cluster analysis for portfolio optimization, *Journal of Economic Dynamics and Control*, 32 (1) 235-258.

[6] León, D., Aragón, A., Sandoval, J., Hernández, G., Arévalo, A., Nifio, J. (2017). Clustering algorithms for Risk-Adjusted Portfolio Construction, *Procedia Computer Science*, vol. 108, p. 1334- 1343. Available: 10.1016/j.procs.2017.05.185.

[7] Jain, A. (2010). Data clustering: 50 years beyond *K*-means, *Pattern Recognition Letters*, 31 (8) 651-666. Available: 10.1016/j.patrec.2009.09.011.

[8] Singh, A., Yadav, A., Rana, A. (2013). K-means with Three different Distance Metrics, *International Journal of Computer Applications*, 67 (10) 13-17. Available: 10.5120/11430-6785.

[9] Fu, T. (2011). A review on time series data mining, *Engineering Applications of Artificial Intelligence*, 24 (1) 164-181. Available: 10.1016/j.engappai.2010.09.007.

[10] Berndt, D. J., Clifford, J. (1994). Using dynamic time warping to find patterns in time series. *In*: KDD workshop, 10 (16) 359-370.

[11] Kruskal, J. B., Liberman, M. (1999). The symmetric time-warping problem: from continuous to discrete, *In*: Kruskal J. B., Sankoff D., editors. Time Warps, String Edits, and Macromolecules: The Theory and Practice of Sequence Comparison. Stanford: CSLI Publications; p. 125-161.

[12] Keating, C., Shadwick, W. F. (2002). A universal performance measure, *Journal of Performance Measurement*, 6 (3) 59-84.

[13] Calinski, T., Harabasz, J. (1974). A dendrite method for cluster analysis, *Communications in Statistics - Theory and Methods*, 3 (1) 1-27. Available: 10.1080/03610927408827101.