Eurasian Journal of Economics and Finance, 6(3), 2018, 1-12
DOI: 10.15604/ejef.2018.06.03.001

EURASIAN JOURNAL OF ECONOMICS AND FINANCE
www.eurasianpublications.com

THE CLUSTER ANALYSIS IN THE MANUFACTURING INDUSTRY WITH K-MEANS METHOD: AN APPLICATION FOR TURKEY

Zeynep Karaca
Erzurum Technical University, Turkey
Email: zeynep.karaca@erzurum.edu.tr

Abstract

Clusters are networks of firms and manufacturers that are strongly connected within the production chain, creating value added together. There are some demand and supply-side advantages of being in a cluster, including proximity to the consumer, reduction in the cost of finding customers, reputation, information externality, output competition, knowledge distribution, an expert workforce, and infrastructural utility. With clusters, employees can be prevented from being employed in different places and skilled workers can be attracted. In addition, clustering facilitates access to key inputs and reduces transaction costs through the use of local suppliers, meaning that all market, technology and competitive information accumulates in the cluster. This study investigates data mining in Turkey using cluster analysis techniques to determine how the manufacturing of wood products industry sectors are clustered among the regions of Nace Rev. This is done using the number of local units and the number of employees in the manufacturing industry sectors according to two classifications. Data obtained from this study will be used in the Annual Industry and Service Statistics published by Turkey's Statistics Institute. Since the most recently published data was provided in 2015, the number of local units and employment figures for 2015 will be used in the study. The algorithm for the k-means method used in this study was written using Microsoft Excel.

Keywords: Cluster, K-Means, Algorithm

JEL Classifications: B21, C15

1. Introduction

Clustering studies have become a focus of attention, especially in the case of Porter (1990), who conducted a clustering study of ten countries to analyze the forces behind the competitive advantages of each nation. Clustering is defined as a group of firms, information and innovation centers that are functionally close to each other to discover new market strategies, products or new methods. Most clustering initiatives aim to connect and communicate between companies operating in the same geographical area or business area. A cluster is also an economical network of independent firms, information centers, or consumers (Kristensen and Laursen, 1999).

Clusters have different definitions and were initially defined as sectors that were related to each other through official production associations without regard to geographic proximity. Later, when these clusters begin to show high geographical accumulation, they are defined as industrial complexes (Feser and Bergman, 2000).

Porter (1998) has made a cluster definition, which covers most of the others, as an alternative way of organizing the supply chain. Compared to the stock market, where there are scattered and random buyers and sellers, the geographical proximity of firms and industries in a region and the repeated transactions between them provides a better-coordinated and more trusting environment. A cluster of activities, both independent and informally linked to firms and institutes, creates a powerful organization that provides efficiency and flexibility (Porter, 1998: 79–80). Although there is no definitive cluster description, the definition of a cluster of players who are directly or indirectly active or have the potential to produce a product or service is accepted as commonly geographically concentrated. In the light of this definition, clustering can be defined as a concentration of geographical regions of interdependent firms. Clusters include firms that produce similar products with common inputs, such as technology and the labor force, both vertically and horizontally. Many clusters include state institutions and universities, agencies, vocational training institutions, and trade associations (Eraslan, 2008).

The basic characteristics of a cluster can be listed as follows (Boronenko and Zeibote, 2011):

. The cluster is not an entity, but an economic entity,

. Participants of the conglomerate are economically dependent on each other even though they are legally independent,

. Participants of the conglomerate differ in economic status and type of activity,

. Participants of the co-op are geographically close together and work in the same region.

It is possible to talk about the existence of three kinds of clusters (Seçilmiş, 2015): Pure cluster model, based on geographic proximity, draws attention to the declining transaction costs of the cluster, the effects on the specialized local workforce, information flow, and externality. Industrial complex model uses identifiable and stable relationships between companies based on commercial ties to identify the cluster. Social network model not official, but it is worth mentioning as it focuses on the social integration of roles and economic activities of institutions and networks.

The companies in the cluster establish good network communications and links that provide information and technology externalities. Closer relationships and information externalities with customers and other companies encourage manufacturers to develop new ideas because it reduces the cost of the experience of cluster formation (Sosnovskikh, 2017). The activities of firms in one sector can influence the performance of producers in other regional sectors (Feser and Bergman, 2000) through the development and diffusion of new technologies, the efforts of producers to use advanced production techniques, and the exchange of knowledge and skills. Clusters are the geographical combinations of interconnected companies, specialized producers, service providers, companies in related industries and partnerships (such as universities and trade associations) that compete in a particular area. Clustering is an integral part of the innovation process (Haviernikova, 2012).

The cluster is an environment where many businesses will co-operate to gain different economic advantages. The Hollywood and Bollywood film industry, California wine industry, Silicon Valley and Boston are examples of globally known clustering of information technology (Boja, 2011). The wine cluster diagram from Porter's (1998) clusters is shown in Figure 1.

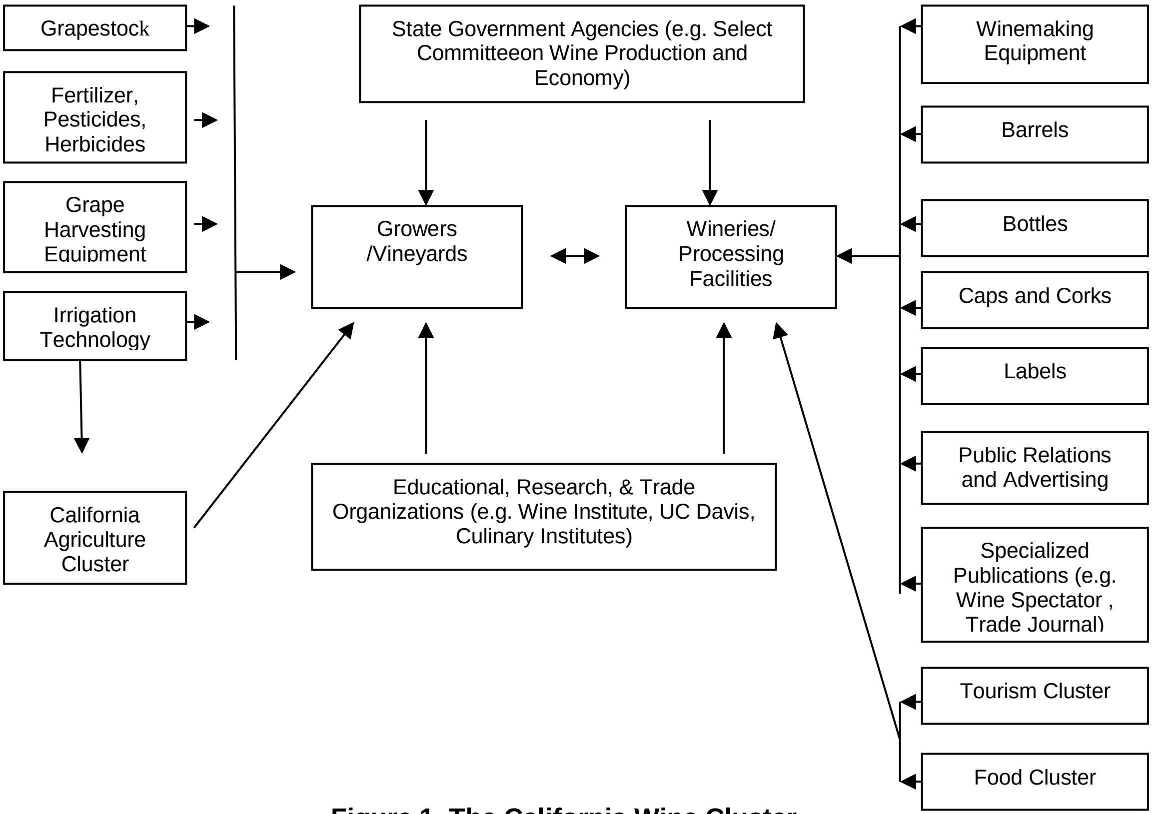

Figure 1. The California Wine Cluster
Source: Porter (1998)

As shown in Figure 1, the wine cluster in California contains public and university elements, such as profitable wines, grape producers, complementary companies that provide equipment for production, supporting advertising and corporate communication units, the University of California wineries program, and committees in the senate.

The various advantages of the cluster are as follows:

- The gathering of companies creates a huge market, which allows them to reach more consumers.
- Transport and supply chain costs are reduced.
- Resource access is cheaper.
- Specialization of products and services increases.
- A more competitive environment that provides better motivation emerges.
- There is greater cooperation between cluster members and closeness increases trust among firms and facilitates communication.
- Better access to a qualified workforce develops (Boja, 2011).

Some studies used an input–output approach to examine the determination of clusters. The main objective of the input–output model was to model the relationship between industrial sectors in an economic system and elaborate on the many relationships between these sectors. Morrissey and O'Donoghue (2013) used an input–output approach to examine the maritime transport industry and the potential cluster formation in this sector. Stimson *et al.* (2006) developed an approach to measure industry loyalty using an input–output approach to understand the strengths of an industry cluster. Feser and Bergman (2000) extracted a pattern of the clustering of detailed sectors using links in the input–output table and this template has been applied to the manufacturing industry in the North Carolina region; they stated that this template could be further developed and applied to a group of companies that are related to one another.

Florea (2015) analyzed what the determinant factors that influence companies to become a member of a cluster. For this, surveys were conducted to collect all data by selecting an example from the participants. Sosnovskikh (2017) conducted a cluster analysis for Russia, which he obtained through face-to-face meetings with managers, donors, and representatives. Vertakova *et al.* (2016) also developed a classification methodology to investigate both innovative industries in the Russian economy and the impact of these industries on the gross domestic product. For this, they first classified the innovation activities of the manufacturing industry and then noted the impact of the innovation process on the industry.

Some researchers state that clusters are spontaneous (Brusco, 1982; Delgado *et al.* 2014), meaning that communication between the markets can lead to the formation of spontaneous clusters by combining a talented workforce with social and institutional capital. Some researchers emphasized the role of national and international public policies (Cowling and Sugden, 1999; Parrilli, 2009). In particular, they argued that the cluster formation was not spontaneous, that instead of implementing national law and initiatives it will allow cluster formation. Ellison *et al.* (2010) noted that proximity to consumers and producers is one of the most important reasons for cluster formation. Yang *et al.* (2012) identified clusters using registered data from businesses and tried to show how they changed over time. In addition to traditional production factors that are widespread among researchers, knowledge and skill are influential in cluster formation and increasingly important factors.

Sutikno (2015) used the location quotient (LQ) method to investigate clusters in the manufacturing industry, and Yuxiang *et al.* (2011) used the LQ method to determine industries' specialization ratios and clusters. It has been determined that the development of industrial clusters increases competition and provides economic development. Reveiu and Dardala (2013) used the LQ method to determine clusters in Romania. There is a relationship between the results obtained and university research in the regions that contain these clusters. Accordingly, there is a strong relationship between clustering and the scientific studies conducted by academic staff in the regions where these clusters are located; these research areas can provide important information about cluster types.

Jan *et al.* (2012) built a dynamic model to conduct a deep analysis of industry clusters in three respects (talent, technology, and capital) to examine software industry clusters in China. Lin, Tung, Huang (2006) used a system dynamics approach to determine the factors affecting the formation of industrial clusters. Researchers have established a dynamic model to identify the factors affecting industrial clustering and have indicated that this approach can be used to analyze mixed relationships between factors affecting cluster formation.

Jarungkitkul and Sukcharoensin (2016) used the Porter Diamond Model to determine the competition between clusters, and Chung (2016) used the Porter Diamond Model to investigate the competitiveness of logistics clusters. Various results have been achieved by comparing the competition of logistic clusters among Asian countries.

Clusters are present in studies on the variables that affect them. Clustering studies at the regional level examine the clustering functions of a single industry and some researchers have adopted macro-level studies. For example, Bathelt and Li (2014) showed that Chinese and Canadian clusters are directly related to foreign investments. Li (2014) showed that trade relations will develop clusters. Lu *et al.* (2016) adopted that microeconomics cannot replace macroeconomics and that the inspections of individual firms in clusters are inadequate. To gain perspective on regional policy makers in China, both the relationships between clusters and how clusters will affect co-located clusters has been researched. Panel data was used to determine how clusters would affect the aggregate factor productivity of other co-located clusters.

In addition to these studies, Bzhalava and Cantner (2018) and Zoia *et al.* (2018) has found that the channels of cooperation increase firms' productivity and R & D investments.

The study consists of three parts. The first is a summary of the literature that details the importance of clustering and why it should be done. The second gives more detailed information about clustering; for this, the Diamond Model of Michael Porter—the founder of clustering theory—is described. Porter (1990) described the diamond model in which an entity develops successes in certain industries and failures in others. According to Porter (1990), success is achieved in business clusters. These clusters are industrial areas in which enterprises and various public and private sector institutions operate together with suppliers and other connections. The emergence of these clusters will increase the speed with which competitive advantages are obtained in countries and industries (Porter, 1990).

The third part provides clustering methods that include data mining models with classification, clustering, and association rules. Clustering is the process of separating classes and clusters, and clustering algorithms commonly used in the literature include partitioning methods, hierarchical methods, density-based methods, grid-based methods, and model-based methods. This study performs clustering analysis with partition-based methods. The most common partitioning methods are the k-means and k-medoid methods. In partitioning methods, n is the number of objects in the database and k is the number of sets to be created. The partitioning algorithm divides n objects into k coils; when objects in the same cluster are similar to one another, they are different from the objects in different coils (Ozekes, 2003).

The fourth part gives the results of clustering analysis using the k-means method. According to the results obtained after the K-means algorithm has been executed, information will be obtained on how firms that operate in the manufacturing industry should cluster.

2. The Clustering

Clustering is an approach that enhances national and regional competitiveness. A cluster-focused approach emphasizes externalities that are very important in terms of competition; it draws attention to the interrelationships and network structures that firms have with each other and with the public, academia, research, and supporting institutions. Industrial clusters have three defining characteristics, the first of which is closeness; clusters facilitate regional accessibility, which is important. The second property is value creation; aggregates include firms that are related to each other through the production of goods and services. The third characteristic is the business environment; firms share a business environment created by individual activities, cooperation, the public, universities, and other organizations. The balance between co-operation and competition in the workplace is an important determinant of industrial clusters' innovative capacities (Sosnovskikh, 2017).

The objectives of the cluster are increasing competitiveness, innovation, and productivity, increasing exports, creating an industry structure with a qualified labor force. Research centers, incubation centers, technology transfer, innovation centers, and technology development zones should focus on specific areas, be integrated in an integrated manner, and encourage them to support their respective cluster activities. For this, innovative and high value-added cluster formation will be encouraged by considering value chain relationships at the regional level, and cooperation among enterprises in existing clusters will be increased.

Industry clustering is defined as firms operating in the same area or in the same geographical region. Firms in an industry cluster have a competitive advantage in the industry and encourage innovation (Jan *et al.* 2012). Competitiveness may be improved when cluster activities emerge in the same or similar areas. These firms are economically interconnected, and share a common infrastructure and industrial environment (Yang *et al.* 2012). Industrial clustering policy is a widely used tool in economic development planning; it is believed that industrial clusters, which are groups of geographically close companies within the same industry, can increase employment, diversify exports, and lead to technology transfer. The key elements of the industrial cluster model are strategic programs and state regulation (Sosnovskikh, 2017) that allow for a cooperative, competitive environment, appropriate geographic location, proximity to resources, and innovation and productivity. Research on social and economic development depends on substantial development in the cluster. An effective clustering policy enhances competitiveness and innovation capacity. Therefore, clustering policy is the most innovative and focuses on industries with high growth potential (Vertakova *et al.* 2016). Some industry clusters co-create employment and improve welfare (Lu *et al.* 2016); industrial clustering improves expertise and strengthens industry cohesion. This industry model encourages the rational distribution of technology, and both ability and capital, and can successfully manage the innovation process. Moreover, industrial clustering is highly influential on regional industrial order and the optimization of regional economic structure (Yuxiang *et al.* 2011).

Clusters are an increasing force of exports and a driving force with which to attract foreign investment. In addition, clusters can lead to the creation of new dialogue between institutions such as government institutions, universities, and schools. Clusters can take different forms depending on their depth and complexity; these can include local units such as restaurants, car dealers, antique dealers, and clustered large and small businesses. The boundaries of clusters change with new firms and industries. Technological advances and the developments of the market reveal new industries and create new connections. Regulatory changes such as telecommunications and transport can broaden clusters' boundaries (Porter, 1998).

The method applied in this study has been used in many other studies; however, these methods are mostly found in studies carried out in the field of natural sciences. For example, studies on data mining are examples in computer engineering. In data mining, clustering algorithms are used to group data with common properties. The study is applied to the field of economics and a k-means clustering method is used in original data mining to ensure that sectors are clustered correctly according to the numbers of local units and employees. In the field of economics, clustering frequently uses interviews, fieldwork, and focus-group techniques. Information requirements and competition conditions are discussed with knowledgeable suppliers, producers, sellers, customers, and public and university personnel. In this study, the application of clustering methods used in the field of computer engineering to the field of industrial clustering is considered an innovation. The study may provide a basis for the application of clustering methods outside the k-means method to this area.

In the K-means algorithm, the k value, which is the number of clusters, needs to be determined first. The algorithm distributes n observations to k clusters and cluster similarity is measured by the mean value of observations in the cluster, which is a cluster's center of gravity. According to the algorithm's operating mechanism, k centers are first chosen randomly, where each represents the center or average of a coop. The remaining observations are included in the closest clusters, taking into account the distances of the clusters relative to the mean values. Then, the average value of each cluster is calculated, new cluster centers are determined, and the observation-center distances are re-examined. The algorithm continues until there the clusters no longer change.

The Classification of Statistical Regional Units has been defined throughout the country to collect and develop regional statistics, to make socio-economic analyzes of the regions, to determine the framework of regional policies and to establish comparable statistics database in accordance with the European Union Regional Statistical System. According to this classification to be statistically Turkey is divided into 26 NUTS II Region. This distinction is not an administrative classification but a statistical classification.

This work's aim is to find out how to correctly cluster the sectors operating in NUTS II Region of the manufacturing industry.

The following steps will be followed to achieve this goal.

1. The Turkey Statistics Level 2 in each region of the Annual Industry and Service Statistics published by the Institution, the number of employees employed in each sector, and the number of local units will be obtained.
2. An algorithm will be written for the k-means method, which is a data mining method.
3. The solution to the written algorithm will determine which clusters need to be set up.

3. Method and the Clustering Results

Data mining can be defined as accessing information between large-scale data and storing information. In other words, data mining is a collection of processes that involve the use of advanced data analysis tools such as statistics, artificial intelligence, and machine learning to reveal hidden patterns and associations within the data that do not stand alone. Data mining allows you to obtain critical information from very large datasets; this information is used to make objective assessments and strategic decisions. The computer is responsible for determining the data's relationships, rules, and specifications and the goal here is to be able to detect previously unrecognized data patterns. Clustering enters data-mining techniques into descriptive (i.e.

unsupervised) classification. In surveillance classrooms, the purpose is to classify given initial and unclassified verbs into meaningful sub-sets (Sariman, 2011).

Clustering analysis is a data mining technique that allows objects in databases to be grouped together to bring objects with similar properties together. Clustering analysis has been used in a wide variety of fields including engineering, civil engineering, and agriculture. Clustering is an optimization process, and optimal clustering can be achieved by minimizing the sum of the clusters' mean distances from cluster elements (Colak *et al.* 2016).

Clustering is the grouping of objects according to the cluster's first purpose characteristics; high homogeneity should be exhibited both within and among clusters. Clustering algorithms can be hierarchical and non-hierarchical.

There are two subclasses in the hierarchical class: stacker and divisive approaches. Meanwhile, the non-hierarchical class is divided into four sub-classes: partition, density-based, grid-based, and model-based clustering approaches. Hierarchical algorithms combine two objects that are the most similar to one another in a set, while non-hierarchical clustering algorithms are direct clustering algorithms (Taskin and Emel, 2010). The main difference between these two analyses is that while the number of clusters is determined by analyzing the number of clusters without preliminary knowledge of how many sets of datasets must be sorted in hierarchical clustering methods, it has to be determined how many clusters will be initially clustered in non-hierarchical clustering methods (Akin and Eren, 2012).

Stacker clustering approaches include single link, full link, average link, ward method, and the central method; the only connection method is based on the shortest distance principle. Find two observations that are the closest to each other and set this cluster to the first step, then this core finds another observation close to the cluster or two other observations close to each other to expand the cluster; multiple clusters can form through this method. The only difference between a full connection method and a single connection method is the start point of the two most distant observations. The average connection method is based on observation in the middle of a coop and the Ward method is based on the mean distance from observations in the middle of a coop to observations in the same coop. The central method is based on the average of the observations that make up a cluster; if there is only one observation in a cluster, its value is considered the center (Kangalli *et al.* 2014).

The divisor hierarchical method starts with a large cluster of all observations; smaller observations are created by eliminating similar observations. Each observation is continued until the cluster is formed alone (Celik, 2013).

Partition-type clustering approach is the most widely used approach in the non-hierarchical clustering class. Such algorithms usually change the center of the clusters until all points are at their minimum distance to the relevant cluster centers; the most common partition approach example is k-means (Taskin and Emel, 2010).

The density-based clustering approach is another non-hierarchical clustering approach. In this technique, a clustering process is performed according to the density of the data from the cluster selection based on the distance. In the density-based clustering technique, clusters are defined in the database as higher density areas (Bircan and Cam, 2016).

Grid-based clustering approach uses grid structures that consist of square cells in the end number to examine the data space. Using this grid structure means that it is independent from the number of objects in the database. The only thing that affects their performance is the number of frames they use. As the number of frames increases, the calculation time increases and the performance decreases. The most important advantage of the grid-based clustering approach is that it can be reached quickly because of its small processing load (Gulce, 2010).

Model-based clustering approach attempts to express a mathematical model for the data at hand; this method assumes that the data is placed in the data space with a logic composed of certain probability theories (Gulce, 2010).

This study uses the k-means method, which is a non-hierarchical clustering method. The K-means algorithm was developed by MacQueen (1967) and is a sharp clustering algorithm that allows each dataset to belong to only one cluster. Meanwhile, it is also a cyclic algorithm that constantly renews clusters and continues until it has reached an optimal solution. The general concept of the k-means algorithm, one of the earliest clustering methods, is to divide a dataset that consists of n data objects into k clusters that are given as input parameters. The aim here is to ensure that the clusters obtained at the end of the compartmentalization process have maximum intra-cluster similarities and minimum inter-cluster similarities. Cluster similarity is measured by the average value of the distances between objects considered as the center of gravity of the cone and other objects in the cluster. The processing steps of the K-means algorithm are as follows:

1. Step: The first cluster centers are determined by the use of one of two different paths. The first of these paths is to select k random points as the number of clusters among objects; the second way is to determine the center points by taking the average of all objects.
2. Step: The distance to each selected object's center points is calculated. According to the results obtained, all objects are placed in the cluster closest to them in k clusters.
3. Step: The new center points of the resulting clusters are replaced by the average values of all objects in that cluster.
4. Step: Repeat Steps 2 and 3 until the center points change.

The K-means algorithm uses three formulas to calculate the center point distance of each object; these are explained below.

The Euclidean distance formula is not affected by the addition of new objects, which may be extraordinary to the clustering analysis. However, the scale between dimensions significantly affects the differences. The Euclidean distance formula is the most widely used distance formula.

The Manhattan distance formula equals the average difference between dimensions. The effect of discomfort decreases when this criterion is used because it is not neglected.

Chebyshev distance formula equals the absolute maximum distance between two objects.

In this study, the Euclidean distance formula will be used to account for the distance.

p= (p1, p2, ………., pn) ve q= (q1, q2, ………., qn)

$$\sqrt{\sum_{i}^{n}(p_i - q_i)^2} = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + ... + (p_n - q_n)^2} \tag{1}$$

The greatest deficiency of the K-means algorithm is that it cannot determine the value of k. Therefore, it is necessary to create experiments for different k values to obtain a successful clustering (Demiralay and Çamurcu, 2005). The most practical way to determine the number of clusters is to use Equation (2).

$$k = \sqrt{\frac{n}{2}} \tag{2}$$

Where k is the number of clusters, n is the number of data to be clustered. The quadratic error criterion (SSE) is the most widely used criterion for evaluating the K-means clustering method. The best result gives clusters with the lowest SSE value (Colak *et al.* 2016).

$$SSE = \sum_{i=1}^{k}\sum_{x \in C_i} dist^2(m_i, x) \tag{3}$$

Where x denotes the center point of the set Ci.

This study uses a k-means algorithm to determine the appropriate cluster numbers according to the number of employees employed in the manufacturing industry and the number of local units. The number of local units and the number of employees will be used for these Annual Industry and Service Statistics released by the Turkey Statistical Institute data. Since this data was last published in 2015, the 2015-year data is taken as the basis for the study. A suitable number of kernels, k will first be determined for the algorithm to be written. Cluster formation will then be established using the Euclidean distance formula. The advantage of the K-means algorithm over other clustering techniques is that it is effective at clustering large datasets. Meanwhile, the k-means algorithm in the clustering of large datasets is faster than in hierarchical clustering algorithms. The only difficulty of this algorithm lies in determining the number of clusters (Huang, 1997). However, the most obvious advantage of the k-means algorithm is its reliability; it is a cyclic algorithm that continuously updates clusters and continues until it reaches the optimal solution; the algorithm repeats itself until the cluster centers become unchanging. A proposal can be made on how to set up the firm cluster operation in the manufacturing industry by identifying the correct clusters and cluster centers; Microsoft Excel can be used to write this algorithm.

The clustering results obtained by the solution of the written algorithm are shown in the following table. Table 1 shows how the sectors operating in the manufacturing industry need to be clustered on sub-regional bases.

Table 1. The Clustering Results

Cluster 1	Cluster 2	Cluster 3
TR10 (İstanbul)	TR82 (Kastamonu, Cankiri, Sinop)	TRA1 (Erzurum, Erzincan, Bayburt)
TR42 (Kocaeli, Sakarya, Duzce, Bolu, Yalova)	TR83 (Samsun, Tokat, Corum, Amasya)	TRA2 (Agri, Kars, Igdir, Ardahan)
	TR90 (Trabzon, Ordu, Giresun, Rize, Artvin, Gumushane)	TRB1 (Malatya, Elazig, Bingol, Tunceli)
		TRB2 (Van, Mus, Bitlis, Hakkari)
		TRC3 (Mardin, Batman, Sirnak, Siirt)

4. Conclusion

It is important how the cluster should be clustered in today's increasingly important world. Clustering is an approach that enhances national and regional competitiveness. A cluster-focused approach emphasizes externalities that are very important in terms of competition; it draws attention to the interrelationships and network structures that firms have with each other and with the public, academia, research, and supporting institutions. Industrial clusters have three defining characteristics, the first of which is closeness; clusters facilitate regional accessibility, which is important. The second property is value creation; aggregates include firms that are related to each other through the production of goods and services. The third characteristic is the business environment; firms share a business environment created by individual activities, cooperation, the public, universities, and other organizations. The balance between co-operation and competition in the workplace is an important determinant of industrial clusters' innovative capacities The objectives of the cluster are increasing competitiveness, innovation, and productivity, increasing exports, creating an industry structure with a qualified labor force. Research centers, incubation centers, technology transfer, innovation centers, and technology development zones should focus on specific areas, be integrated in an integrated manner, and encourage them to support their respective cluster activities. For this, innovative and high value-added cluster formation will be encouraged by considering value chain relationships at the regional level, and cooperation among enterprises in existing clusters will be increased

This study used the k-means clustering algorithm in the field of data mining was applied to the field of economy and tried to determine the correct clusters. According to the results obtained with the solution algorithm, the manufacturing industry in Turkey should consist of three clusters. The application of the k means clustering method, which is mostly used in the field of engineering, to the clustering in the economic field constitutes the most important part of the study.

The k-means method used only two variables since clustering was done according to the X and Y axes. However, a large number of variables can affect cluster formation. In other words, cluster analysis can be performed by considering a large number of variables; this issue should be investigated in future studies.

References

Akin, H. B. and Eren, O., 2012. OECD ulkelerinin egitim gostergelerinin kumeleme analizi ve cok boyutlu olcekleme analizi ile karsilastirmali analizi [OECD countries with education indicators comparative analysis of cluster analysis and multi-dimensional scaling analysis]. *Oneri Dergisi*, 10(37), pp. 175-181.

Bathelt, H., and Li, P. F., 2014. Global cluster networks—Foreign direct investment flows from Canada to China. *Journal of Economic Geography,* 14(1), pp. 45–71. https://doi.org/10.1093/jeg/lbt005

Bircan, H. and Cam, S., 2016. Veri madenciliginde kumeleme analizi ve saglik sektorunde bir uygulamasi [Clustering analysis in data minig and an application in health sector]. *C. Ü. İktisadi ve İdari Bilimler Dergisi*, 17(2), pp. 85–96.

Boja, C., 2011. Clusters models, factors and characteristics. *International Journal of Economic Practices And Theories*, 1(1), pp. 34–43.

Boronenko, V. and Zeibote, Z., 2011. The potential of cluster development and the role of cluster support policies in Latvia. *Economic Annals,* 56(191), pp. 35–67. https://doi.org/10.2298/EKA1191035B

Brusco, S., 1982. The Emilian model: Productive decentralisation and social integration. *Cambridge Journal of Economics*, 6(2), pp. 167–184.

Bzhalava, L., and Cantner, U., 2018. The journey towards open innovation: Why do firms choose different routes? *Eurasian Business Review*, 8(3), pp. 245-265. https://doi.org/10.1007/s40821-017-0101-9

Chung, T. W., 2016. A study on logistics cluster competitiveness among Asia main countries using the Porter's diamond model. *The Asian Journal of Shipping and Logistics*, 32(4), pp. 257–264. https://doi.org/10.1016/j.ajsl.2016.12.010

Cowling, K. and Sugden, R., 1999. The wealth of localities, regions and nations: Developing multinational economies. *New Political Economy*, 4(3), pp. 361–378. https://doi.org/10.1080/13563469908406409

Celik, S., 2013. Kume analizi ile saglik gostergelerine gore Turkiye'deki illerin siniflandirilmasi [Classification of provinces in Turkey according to health indicators by cluster analysis]. *Dogus Universitesi Dergisi*, 14(2), pp. 175–194.

Colak, B., Durdag, Z., and Erdogmus, P., 2016. K-Means algoritmasi ile otomatik kumeleme [Automatic clustering with k-means]. *El-Cezeri Fen ve Mühendislik Dergisi*, 3(2), pp. 315–323.

Delgado, M., Porter, M. E., and Stern, S., 2014. Clusters, convergence, and economic performance. *Research Policy*, 43(10), pp. 1785–1799. https://doi.org/10.1016/j.respol.2014.05.007

Demiralay, M. and Camurcu, A. Y., 2005. Cure, Agnes ve K-Means algoritmalarindaki kumeleme yeteneklerinin karsilastirilmasi [Comparison of clustering characteristics of cure, agnes and k-means algorithms]. *Istanbul Ticaret Universitesi Fen Bilimleri Dergisi,* 4(8), pp. 1–18.

Ellison, G., Glaeser, E., and Kerr, W., 2010. What causes industry agglomeration? Evidence from coagglomeration patterns. *American Economic Review*, 100(3), pp. 195–1213. https://doi.org/10.1257/aer.100.3.1195

Eraslan, İ. H., Bulu, M., and Bakan; I., 2008. Kumelenmeler ve inovasyona etkisi: Turk turizm sektornde uygulamalar [Clusters and their effects on innovations: implementations in the turkish tourism sector]. *Seyahat ve Otel İşletmeciliği Dergisi*, 5(3), pp. 1–35.

Feser, E. J., and Bergman, E. M., 2000. National industry cluster templates: A framework for applied regional cluster analysis. *Regional Studies,* 34(1), pp. 1–19. https://doi.org/10.1080/00343400050005844

Florea, C. A., 2015. Cluters- A strategy for long run development. *Eurasian Journal of Social Sciences*, 3(1), pp. 20-24.

Gulce, A. C., 2010. *Veri madenciliginde Apriori algoritmasi ve Apriori algoritmasinin farkli veri kumelerinde uygulanmasi*. Unpublished Ph.D. Trakya University.

Haviernikova, K., 2012. Selected aspects of clusters initiatives in the field of industry in the Slovak Republic. In Proceedings of 21st International Conference on Metallurgy and Materials (METAL 2012) (pp. 1838–1845).

Huang, Z., 1997. *A fast clustering algorithm to cluster very large categorical data sets in data mining*, Workshop on Research Issues on Data Mining and Knowledge Discovery.

Jan, C.G., Chan, C. C., and Teng, C. H. 2012. The effect of clusters on the development of the software industry in Dalian, China. *Technology in Society*, 34, pp. 163–173. https://doi.org/10.1016/j.techsoc.2012.01.002

Jarungkitkul, W., and Sukcharoensin, S., 2016. Benchmarking the competitiveness of the ASEAN 5 equity markets: An application of Porter's diamond model. *Benchmarking*, 23(5), pp. 1312–1340. https://doi.org/10.1108/BIJ-05-2014-0047

Kangalli, S. G., Uyar, U., and Buyrukoglu, S., 2014. OECD ulkelerinde ekonomik ozgurluk: Bir kumeleme analizi. [Economic freedom in oecd countries: a cluster analysis], *Uluslararası Alanya İşletme Fakültesi Dergisi*, 6(3), pp. 95–109.

Kristensen, I. D. F. S., and Laursen, K., 1999. Cluster studies as a basis for industrial policy: The case of Denmark. *Industry and Innovation*, 6(2), pp. 171–190. https://doi.org/10.1080/13662719900000010

Li, P.-F. 2014. Global temporary networks of clusters: Structures and dynamics of trade fairs in Asian economies. *Journal of Economic Geography,* 14(5), pp. 995–1021. https://doi.org/10.1093/jeg/lbu009

Lin, C., Tung, C., and Huang, C., 2006. Elucidating the industrial cluster effect from a system dynamics perspective. *Technovation,* 26, pp. 473–482. https://doi.org/10.1016/j.technovation.2004.11.008

Lu, R., Ruan, M., Reve, T. 2016. Cluster and co-located cluster effects: An empirical study of six Chinese city regions. *Research Policy*, 45(10), pp. 1984–1995. https://doi.org/10.1016/j.respol.2016.07.003

MacQueen, J. 1967. *Some methods for classification and analysis of multivariate observations*. In Proceedings of the fifth Berkeley symposium on mathematical statistics and probability (Vol. 1, No. 14, pp. 281-297).

Morrissey, K. and O'Donoghue. 2013. The potential for an Irish maritime transportation cluster: An input–output analysis. *Ocean & Coastal Management*, 71, pp. 305–313. https://doi.org/10.1016/j.ocecoaman.2012.11.001

Ozekes, S. 2003. Veri madenciligi modelleri ve uygulama alanlari [Data mining models and application areas]. *İstanbul Ticaret Üniversitesi Dergisi, 2*(3), pp. 65–82.

Parrilli, M. D. 2009. Collective efficiency, policy inducement and social embeddedness: Drivers for the development of industrial districts. *Entrepreneurship Regional Development*, 21(1), pp. 1–24. https://doi.org/10.1080/08985620801886513

Porter, M., E. 1990. *The competitive advantage of nations*. New York: Free Press.

Porter, M. E., 1998. Clusters and the new economics of competition. *Harvard Business Review*, 76, pp. 77–90. https://doi.org/10.1007/978-1-349-11336-1

Reveiu, A. and Dardala, M., 2013. The role of universities in innovative regional clusters. Empirical evidence from Romania. *Social and Behavioral Sciences*, 93, pp. 555–559. https://doi.org/10.1016/j.sbspro.2013.09.238

Sariman, G., 2011. Veri madenciliginde kumeleme teknikleri uzerine bir calisma: K-Means ve K-Medoids kumeleme algortimalarinin karsilastirilmasi [A study of clustering techniques in data mining: comparison of the k-means and k-medoids clustering algorithms]. *Süleyman Demirel Üniversitesi Fen Bilimleri Enstitüsü Dergisi*, 15(3), pp. 192–202.

Secilmiş, İ. E. 2015. Türkiye'de Yaratıcı Endüstrilerin Kümelenmesi [Clustering of creative industries in turkey]. *Ege Akademik Bakış*, 15(1), 9-18.

Sosnovskikh, S., 2017. Industrial clusters in Russia: The development of special economic zones and industrial parks. *Russian Journal of Economics*, 3, pp. 174–199. https://doi.org/10.1016/j.ruje.2017.06.004

Stimson, R., Stough, R., and Roberts, B., 2006. *Regional economic development: Analysis and planning strategy. 2nd ed.* Berlin: Springer.

Sutikno, Muhammad Sri Wahyudi Suliswanto, 2015. The Development of Manufacturing Industry Cluster as an Effort of Economic Improvement Expansion in East Java. *Social And Behavioral Sciences*. 211(25), 992-998. https://doi.org/10.1016/j.sbspro.2015.11.132

Taskin, C. and Emel, G. G., 2010. Veri madenciliginde kumeleme yaklasimlar ve Kohonen aglari ile perakendecilik sektorunde bir uygulama [Clustering approaches in data mining and an application with kohonen networks in retailing sector]. *Süleyman Demirel Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi,* 15(3), pp. 395–409.

Vertakova, Y., Grechenyuk, O., and Grechenyuk, A., 2016. Identification of Clustered Points of Growth by Analyzing the Innovation Development of Industry, *Procedia Economics and Finance*, 39, 147–155. https://doi.org/10.1016/S2212-5671(16)30264-7

Yang, Z., Sliuzas, R., Cai, J., and Ottens, H. F. L., 2012. Exploring spatial evolution of economic clusters: A case study of Beijing. *International Journal of Applied Earth Observation and Geoinformation*, 19, pp. 252–265. https://doi.org/10.1016/j.jag.2012.05.017

Yuxiang, Z., Xilai, Z., Chun, L., Feng, Y., and Hongyu, W. 2011. The development strategy for industrial clusters in Qingdao. *Energy Procedia*, 5, pp. 1355–1359. https://doi.org/10.1016/j.egypro.2011.03.234

Zoia, M. G., Barbieri, L., Cortelezzi, F., and Marseguerra, G. 2018. The determinants of Italian firms' technological competencies and capabilities. *Eurasian Business Review* (Accepted for publication January 2018) https://doi.org/10.1007/s40821-018-0103-2