Avrupa Bilim ve Teknoloji Dergisi
Özel Sayı, S. 478-484, Ekim 2019
© Telif hakkı EJOSAT'a aittir
Araştırma Makalesi

www.ejosat.com ISSN:2148-2683

European Journal of Science and Technology
Special Issue, pp. 478-484, October 2019
Copyright © 2019 EJOSAT
Research Article

Analysis of Agricultural Credit Performance of Turkey using K-means Clustering Algorithm

Zeynep Ceylan[1*], Selin Sabuncu[2]

[1] Samsun University, Engineering Faculty, Industrial Engineering, Samsun, Turkey (ORCID: 0000-0002-3006-9768)
[2] Ondokuz Mayıs University, Engineering Faculty, Industrial Engineering, Samsun, Turkey (ORCID: 0000-0001-9144-1642)

(Bu yayın International Congress on Human-Computer Interaction,Optimization and Robotic Application kongresinde sözlü olarak sunulmuştur.)

(First received 1 August 2019 and in final form 25 October 2019)

(**DOI:** 10.31590/ejosat.638434)

ATIF/REFERENCE: Ceylan, Z. & Sabuncu, S. (2019). Analysis of Agricultural Credit Performance of Turkey using K-means Clustering Algorithm. *European Journal of Science and Technology*, (Special Issue), 478-484.

Abstract

Agriculture is a significant sector that supplies raw materials to many sectors as well as providing nutrients to humans and animals and ensures employment. The economic crises, rapid population growth, the rise in demand for food products have increased importance and necessity of agriculture. For this reason, agriculture must be supported in order not to be affected by adverse conditions and effects. Thus, agricultural credit is an important factor in the development of the production and investment structure of the agricultural sector. In this study, agricultural credit performance of 81 provinces in Turkey in 2018 was compared by taking into consideration the value of total agricultural production, total cultivated area and the amount of agricultural credit used. The data used in this study were collected from the Banking Regulation and Supervision Agency (BRSA) and the Turkish Statistical Institute. In order to determine relationships between the 81 provinces of Turkey, one of the nonhierarchical clustering method, i.e. the K-means clustering method was applied using SPSS Clementine data mining software. As a result, the credit performance of provinces was evaluated and similarities and differences were revealed using agricultural production value, total cultivated land, agricultural credit volume data.

Keywords: Agricultural credit performance, Clustering, K-means, Data mining

Türkiye'nin Tarımsal Kredi Performansının K-ortalamalar Kümeleme Algoritması ile Analizi

Öz

Tarım, insanlara ve hayvanlara besin sağlamanın yanısıra birçok sektöre hammadde ve istihdam sağlayan önemli bir sektördür. Ekonomik krizler, hızlı nüfus artışı, gıda ürünlerine olan talebin artması, tarımın önemini ve gerekliliğini arttırmıştır. Bu nedenle, olumsuz koşullardan ve etkilerden etkilenmemesi için tarım desteklenmelidir. Dolayısıyla, tarımsal kredi, tarım sektörünün üretim ve yatırım yapısının gelişiminde önemli bir faktördür.
Bu çalışmada, 2018 yılında Türkiye'deki 81 ilin tarımsal kredi performansı, toplam tarımsal üretim değeri, toplam ekili alan ve kullanılan tarımsal kredi miktarı dikkate alınarak karşılaştırılmıştır. Bu çalışmada kullanılan veriler Bankacılık Düzenleme ve Denetleme Kurumu'ndan (BDDK) ve Türkiye İstatistik Kurumu'ndan toplanmıştır. Türkiye'nin 81 ili arasındaki ilişkileri belirlemek için, hiyerarşik olmayan kümeleme yöntemlerinden biri olan K-ortalamalar kümeleme yöntemi, SPSS Clementine veri madenciliği yazılımı kullanılarak uygulanmıştır. Sonuç olarak, tarımsal üretim değeri, toplam ekili alan, tarımsal kredi hacmi verileri kullanılarak illerin kredi performansı değerlendirilmiş ve benzerlikler ve farklılıklar ortaya konulmuştur.

* Corresponding Author: Samsun University, Engineering Faculty, Industrial Engineering, Samsun, Turkey, ORCID: 0000-0002-3006-9768, zeynep.ceylan@samsun.edu.tr

Anahtar Kelimeler: Tarımsal kredi performansı, Kümeleme, K-ortalamalar, Veri madenciliği

1. Introduction

Agriculture is a crucial sector for Turkey in terms of social and economic aspects. The importance of agriculture in the economy of Turkey and other countries is measured as the added-value of the agricultural sector as percent of Gross Domestic Product (GDP). After the 1980s, the share of agriculture in GDP has declined in Turkey as a result of greater emphasis on industrialization, the reduction of the state's tendency for agricultural protection with the laws, and customizations processes. While the contribution of agriculture to GDP was %17,7 in 1987, it has dropped to %6,1 (52,3 billion dollars) in 2016. As shown Fig. 1, with the economic development, the share of the industry and services sector in the economy increases while the share of the agricultural sector in the economy is gradually decreasing.

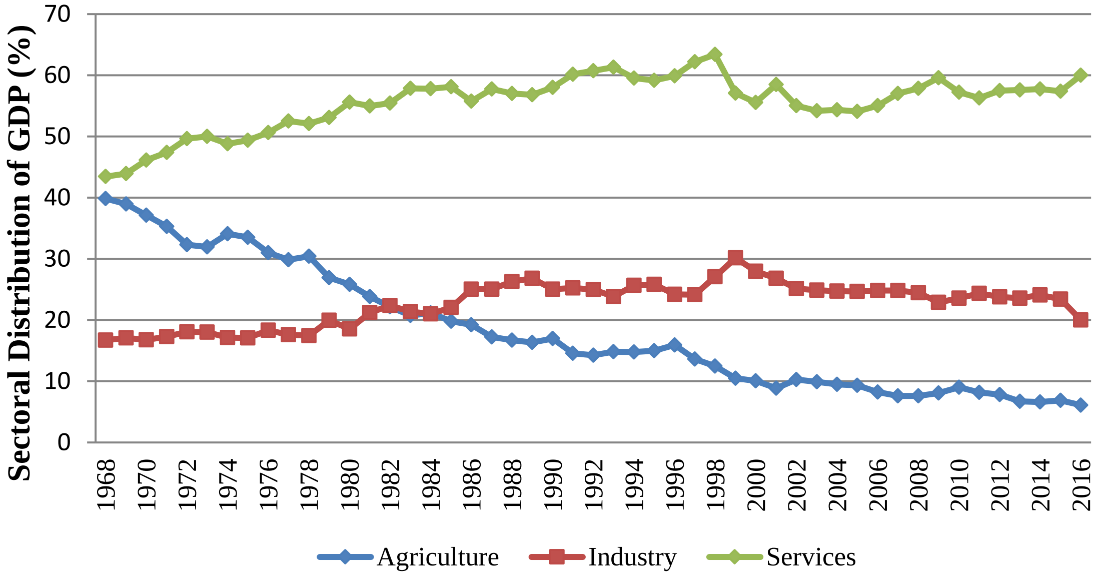

Figure 1. Sectoral Distribution of GDP between 1968-2016

One of the most important factors for the sustainability of production in agriculture is financing (Terin et al., 2014; Adanacıoğlu et al., 2017, Chandio et al., 2017). Because, the credit and input support provided to the agricultural sector directly affects agricultural production. Agricultural credit allows farmers to access new technologies and take advantage of new economic opportunities to increase production and income (e Saqib et al., 2016). For this reason, it must be supported to create a sustainable, competitive and organized agricultural sector (Hayran and Gül, 2018).

Agricultural credit usage in Turkey has experienced a rapid decline between 1997 and 2002. As shown Figure, the use of agricultural credits, which increased to 42.1 billion TL in 1997, declined to 6.6 billion TL in 2002. However, after 2002, the use of agricultural credits started to increase with the improvement of the credit utilization conditions and the provision of credit to the agricultural sector of private banks. The use of agricultural credit, which was 6.5 billion TL in 2002, has increased to 24.8 billion TL in 2012. The agriculture received about 3.3% (68.239 Thousand TL) of total credit used in Turkey, 2017. This can be explained by the fact that agricultural producers receive lower amounts of credit from producers in other sectors.

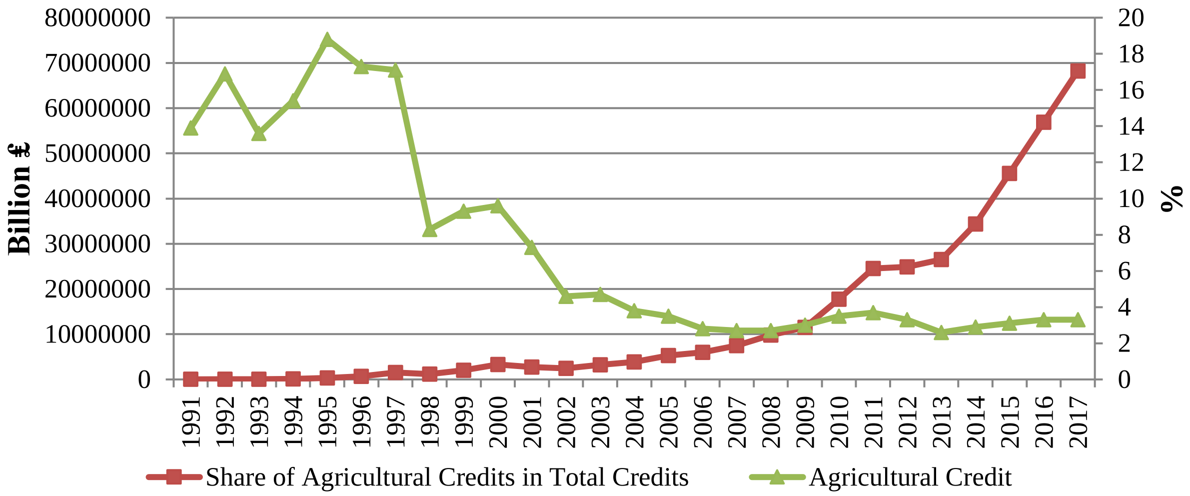

Figure 2. Agricultural Credit Usage in Turkey and Share in Total Loans (%)

In this study, agricultural credit performance of Turkey in 2018 was analzed. The 81 provinces of Turkey were taken as objects of research and they were described by three attributes (variables), i.e. the total agricultural production value, total agricultural credit volume and total cultivated land. In the analysis, k-means clustering method, which is one of the non-hierarchical cluster analysis methods, was used. The objective of the analysis is to group the provinces into clusters (subgroups) that are most similar to each other in the same cluster and most unlike in other clusters. The analyses were carried out based on reports and statistics of The Banks Association of Turkey and Banking Regulation and Supervision Agency (BRSA) and Turkish Statistical Institute (TURKSTAT).

2. Material and Method

2.1. Dataset

The agricultural credit performance of 81 provinces of Turkey in the year 2018 was analyzed according to based on the total agricultural production value (crop and animal production, and live animals values), total cultivated land, and total credit volume variables. Then, per capita values for all variables were obtained by dividing each province by the number of population in 2018. As shown in Table 1, the data were collected from different data sources. The used dataset for the analysis was presented in Table 2.

Table 1. Definition of the Variables

Variable	Definition	Unit of Measurement	Data Source
APV	Agricultural Production Value	Thousand ₺ (TL)	Turkstat*
TCL	Total Cultivated Land	Hectare	Turkstat
TCV	Total Credit Volume	Thousand ₺	BRSA**
POP	Population size	Million	Turkstat

** Turkish Statistical Institute,*
*** Banking Regulation and Supervision Agency*

2.2. Data Mining

Today, rapid development in data collection and storage technology allows organizations to accumulate large amounts of data. However, it is difficult to extract useful information from these data. Traditional data analysis tools and techniques are insufficient to analyze such a large data set. At this point, data mining fills a significant gap. Data mining is a technology that blends algorithms developed to analyze large volumes of data (Kamber and Pei, 2006). Chen et al. (2006) defined data mining as the process of extracting information or patterns from interesting (non-trivial, confidential, previously unknown and potentially useful) information in large database.

2.3. Clustering Analysis

Clustering analysis is one of the important descriptive models used in data mining. Clustering analysis is a statistical technique used in many fields; including image processing, market research, information retrieval, bioinformatics, machine learning, pattern recognition, computer graphics and etc. The main task of cluster analysis is to identify subgroups in the data such that data points in the same subgroup (cluster) are very similar while data points in different clusters are very different.

Several clustering algorithms have been proposed in data mining. In the literature, these algorithms are generally classified as hierarchical and non-hierarchical. In the hierarchical clustering algorithm, two methods are used: agglomerative and divisive. In the agglomerative method, each object is initially considered as a seperate cluster (leaf). Then, the two closest (most similar) clusters are then merged into a new cluster. The process repeats until all data points (clusters) will be merged into the same cluster. In this way, the number of clusters is reduced one step at a time. The resulting cluster structure can be represented by "dendrogram" or tree graph. The most popular algorithms are median method, single linkage, furthest neighbor (complete linkage), centroid method or Ward's method. Each of these algorithms can give different results. On the other hand, the process of the divisive method is the inverse of the agglomerative method. Initially, all the observations are proceed as a single cluster, then the following divisions continue until n clusters are obtained. The main drawback of hierarchical methods is that it is difficult to decide the appropriate cluster number, although there are some suggestions for solving the problem.

2.4. K-Means Method

K-means clustering is one of the most popular and effective non-hierarchical method. The goal of this algorithm is to partition n observations into k clusters. Algorithm is strongly dependent on the value of k. The algorithm runs in iterative (repetitive) steps to define each data point to one of k groups dependin on the provided features. Data points are clustered according to feature similarity.

In K-means method, determining the best number of clusters is a fundamental problem. Because it has a deterministic effect on the clustering results. The amount of clusters must be large enough to represent certain characteristics of the data set. Additionally, the value of k must be less than the number of objects in the data set. In most studies, there are suggestions for determining the appropriate values of k. (Pham et al., 2005).

Table 2. APV, TCV, TCL, and Population Values in the 81 Provinces of Turkey (2018)

Region Code	Province	APV (1000 ₺)	TCV (1000 ₺)	TCL (HA)	Population
TR621	Adana	9755979	4277357	411894	2220125
TRC12	Adiyaman	3151019	641200	175902	624513
TR332	Afyonkarahisar	6839585	1587937	447391	725568
TRA21	Agri	4022055	340480	353496	539657
TR712	Aksaray	4821250	946096	387423	412172
TR834	Amasya	3695394	737720	220755	337508
TR510	Ankara	9637745	8813872	1159710	5503985
TR611	Antalya	13401717	4254141	283203	2426356
TRA24	Ardahan	2248035	526200	38890	98907
TR905	Artvin	1389716	172300	10615	174010
TR321	Aydin	7287372	2894304	141096	1097746
TR221	Balikesir	8544157	2332854	298808	1226575
TR813	Bartin	701590	85804	31441	198999
TRC32	Batman	2094114	256641	81809	599103
TRA13	Bayburt	859069	193465	97069	82274
TR413	Bilecik	1155918	247227	68472	223448
TRB13	Bingöl	1679897	177925	26978	281205
TRB23	Bitlis	1876283	342060	119039	349396
TR424	Bolu	1625705	574052	110256	311810
TR613	Burdur	3570574	728776	140425	269926
TR411	Bursa	8473551	2758012	211661	2994521
TR222	Çanakkale	6032687	1174191	234642	540662
TR822	Çankiri	1729994	563953	201763	216362
TR833	Çorum	4034563	964320	512695	536483
TR322	Denizli	6614214	1873065	271414	1027782
TRC22	Diyarbakir	8817611	1521821	547715	1732396
TR423	Düzce	1151384	429072	11106	387844
TR212	Edirne	3745960	1396325	310021	411528
TRB12	Elazig	3328820	541174	158594	595638
TRA12	Erzincan	2001092	432816	122103	236034
TRA11	Erzurum	6439752	1126376	337992	767848
TR412	Eskisehir	3811756	1089750	537266	871187
TRC11	Gaziantep	7286341	2281402	135181	2028563
TR903	Giresun	1837296	418920	37053	453912
TR906	Gümüshane	947586	201193	81001	162748
TRB24	Hakkâri	1137413	111532	36348	286470
TR631	Hatay	4766449	1441802	138073	1609856
TRA23	Igdir	2460457	399937	97033	197456
TR612	Isparta	3709553	738758	161453	441412
TR100	İstanbul	1695212	4457995	69613	15067724
TR310	İzmir	13761983	6135855	176069	4320519
TR632	Kahramanmaraş	4857753	975539	301698	1144851
TR812	Karabük	415729	104228	33385	248014
TR522	Karaman	4415707	630791	298756	251913
TRA22	Kars	4278282	732509	207446	288878
TR821	Kastamonu	2984027	649898	127034	383373
TR721	Kayseri	5688490	1746937	558247	1389680
TRC13	Kilis	910854	189657	48105	142541
TR711	Kirikkale	1462183	406306	298777	286602
TR213	Kirklareli	2810644	1156418	234323	360860
TR715	Kirsehir	3432697	1104002	347344	241868
TR421	Kocaeli	1547442	651443	66148	1906391
TR521	Konya	19374921	4984523	1837344	2205609
TR333	Kütahya	3136613	588125	306707	577941
TRB11	Malatya	3478114	1001356	187408	797036
TR331	Manisa	8212299	3383677	284905	1429643
TRC31	Mardin	3618128	699356	271442	829195
TR622	Mersin	10952655	2842088	200233	1814468
TR323	Muğla	5756231	2742734	109626	967487
TRB22	Muş	3781157	406052	241553	407992
TR714	Nevşehir	2230745	565501	304199	298339
TR713	Niğde	5002262	842356	239623	364707
TR902	Ordu	3775260	785745	25053	771932
TR633	Osmaniye	2132862	507654	103220	534415
TR904	Rize	2817733	371979	542	348608
TR422	Sakarya	3211528	1067573	78850	1010700
TR831	Samsun	5605900	1396633	257561	1335716
TRC21	Şanlıurfa	12335460	2225012	896645	2035809
TRC34	Siirt	1995328	197831	57665	331670
TR823	Sinop	1039672	223205	73743	219733
TRC33	Şırnak	1482552	115836	102052	524190
TR722	Sivas	5041851	1243149	775370	646608
TR211	Tekirdağ	3696738	1646096	385528	1029927
TR832	Tokat	4983248	816339	289562	612646
TR901	Trabzon	2537680	470222	14356	807903
TRB14	Tunceli	674203	158134	48451	88198
TR334	Uşak	2466370	547667	206183	367514
TRB21	Van	4116608	564375	284239	1123784
TR425	Yalova	396541	192574	5893	262234
TR723	Yozgat	4104498	1223651	609314	424981
TR811	Zonguldak	1082309	163438	25561	599698

3. Results ant Discussion

3.1. Determining Number of Clusters

In order to ensure comparability between the collected data, normalization standardization method was applied by following Eq. 1 (Nisbet et al., 2009).

$$z_i = \frac{x_{i-\bar{x}}}{s_x} \qquad (1)$$

where x and S_x are mean and standard deviation of the used variable in the data, respectively. Using SPSS Clementine software the groupings of provinces was performed with K-means clustering algorithm. In this method, the number of clusters must be determined before starting the clustering analysis. In order to evaluate the k values in the analysis, the preliminary computations were made by agglomerative method (Herbin et al., 2001). Some confirmation of appropriateness of such a number of clusters is the use of a rule of thumb, $k \cong \sqrt{n/2}$, whereby $k \cong \sqrt{81/2}$ (Kijewska and Bluszcz, 2016). Same calculations were performed with the k-means method to confirm this value, assuming k=2,3,4,5 and 6. Table 3 shows the values for the sum of squares error (SSE) for each number of clusters. As a result, the number of cluster with the minimum SSE was specified as 5.

Table 3. Number of Clusters and Sum of Squares Error for K-means Method

Number of Cluster	K=2	K=3	K=4	K=5	K=6
SSE	0.596	0.663	0.755	0.318	0.321

3.2. Clustering by K-means Method

Clustering of provinces in terms of per capita TPV, TCL and TCV shows significant differences in the obtained clusters. Cluster no. 1 comprises of 48 provinces (including İstanbul, İzmir, Ankara and etc), cluster no. 2 and cluster no. 3 have only one province (Kırşehir and Ardahan, respectively), the fourth cluster contains 11 provinces and the fifth cluster consists of 20 provinces. The obtained clusters and corresponding distances from the cluster centers are presented in Table 4.

Table 4. Obtained Clusters with Distances from Centers

Cluster 1				Cluster 2		Cluster 3		Cluster 4		Cluster 5	
Province	Distance	Province	Distance	Province	Distance	Province	Distance	Province	Distance	Province	Distance
Adana	0.162	Kahramanmaraş	0.078	Kırşehir	0.000	Ardahan	0.000	Aksaray	0.125	Afyonkarahisar	0.095
Adıyaman	0.080	Karabük	0.181					Bayburt	0.107	Ağrı	0.250
Ankara	0.153	Kayseri	0.160					Çankırı	0.115	Amasya	0.138
Antalya	0.144	Kilis	0.151					Çorum	0.142	Burdur	0.257
Artvin	0.181	Kocaeli	0.238					Edirne	0.295	Çanakkale	0.142
Aydin	0.320	Malatya	0.051					Karaman	0.376	Erzincan	0.033
Balıkesir	0.201	Manisa	0.257					Kırıkkale	0.256	Erzurum	0.100
Bartın	0.137	Mardin	0.116					Konya	0.149	Eskişehir	0.238
Batman	0.140	Mersin	0.127					Nevşehir	0.117	Gümüşhane	0.183
Bilecik	0.097	Muğla	0.350					Sivas	0.149	Iğdır	0.161
Bingöl	0.129	Ordu	0.104					Yozgat	0.295	Isparta	0.129
Bitlis	0.125	Osmaniye	0.035							Kars	0.317
Bolu	0.194	Rize	0.205							Kastamonu	0.159
Bursa	0.106	Sakarya	0.087							Kırklareli	0.293
Denizli	0.180	Samsun	0.017							Kütahya	0.221
Diyarbakır	0.111	Şanlıurfa	0.199							Muş	0.159
Düzce	0.120	Siirt	0.123							Niğde	0.243
Elâzığ	0.091	Sinop	0.113							Tokat	0.110
Gaziantep	0.084	Şırnak	0.185							Tunceli	0.063
Giresun	0.076	Tekirdağ	0.172							Uşak	0.120
Hakkâri	0.145	Trabzon	0.160								
Hatay	0.097	Van	0.132								
İstanbul	0.273	Yalova	0.180								
İzmir	0.125	Zonguldak	0.219								

The clusters obtained by the K-means method and the effects of variables on clusters are shown in Fig. 3. In the SPSS Clementine, it is accepted that the effects of variables with significance levels below 0.90 on clusters are not significant. As can be seen in Fig. 3, it can be concluded that the effects of all variables on the three clusters are significant.

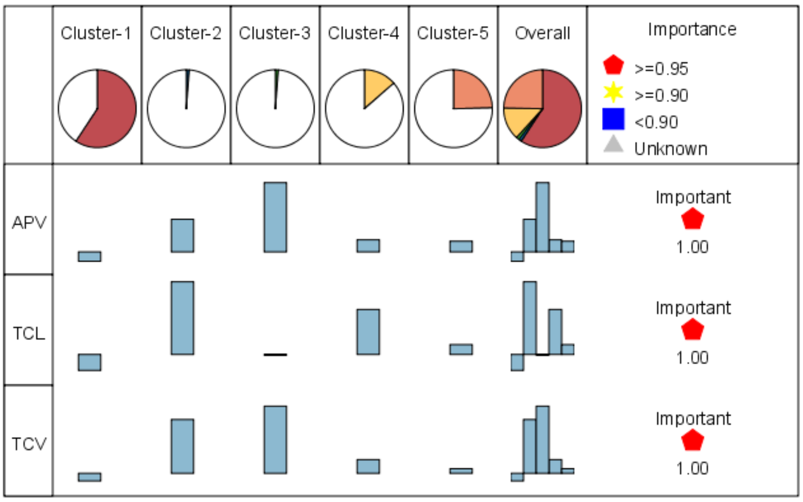

Figure 3. The Clusters Obtained by K-Means Method

3.3. Statistical Analysis

Table 5 shows the average values of used variables per capita in each cluster. Table 6 represents descriptive statistics of clusters on the basis of standardized data for all variables.

Table 5. The Average Values of Variables in Clusters (per Capita)

Cluster No.	APV	TCV	TCL
Cluster 1	4.358	1.099	0.177
Cluster 2	14.192	4.564	1.436
Cluster 3	22.728	5.320	0.393
Cluster 4	9.373	2.302	1.043
Cluster 5	9.071	1.787	0.544

It is seen that 48 provinces creating the first cluster is characterized by the lowest values of average agricultural production value. On the other hand, cluster 2 and cluster 3 have the highest average agricultural production value and total cultivated land. In this case, it can be concluded that the smaller the number of clusters, the higher the average values of agricultural production and cultivated land.

Table 6. Descriptive Statistics of Clusters on the Basis of Standardized Data for Used Variables

Cluster No.	Means	Standart Deviation	Variance
Cluster 1 (48 records)			
APV	*-0.57*	*0.46*	*0.21*
TCL	-0.65	0.35	0.12
TCV	-0.46	0.65	0.42
Cluster 2 (1 record)			
APV	2.00	0.00	0.00
TCL	2.97	0.00	0.00
TCV	3.29	0.00	0.00
Cluster 3 (1 record)			
APV	4.24	0.00	0.00
TCL	-0.03	0.00	0.00
TCV	4.11	0.00	0.00
Cluster 4 (11 records)			
APV	0.74	0.84	0.71
TCL	1.84	0.56	0.31
TCV	0.84	0.59	0.35
Cluster 5 (20 records)			
APV	0.66	0.74	0.55
TCL	0.40	0.30	0.09
TCV	0.28	0.69	0.48

4. Conclusions

The increase in agricultural production, which is one of the basic elements of economic development, depends on the purpose-oriented investments, application of technological innovations and continuous production with increased productivity. One of the most important factors for the continuity of production in agriculture is financing. Agricultural credit is an important factor in the development of the production and investment structure of the agricultural sector in both developed and developing countries and is an important tool of agricultural development. Thus, effective utilization of loans has been one of the main objectives in achieving agricultural development in all development plans and programs.

In this study, the 81 provinces of Turkey in 2018 were divided into clusters using the per capita agricultural production value, agricultural credit volume, and total cultivated area data. For the analysis, the k-means clustering method, which is the most common exploratory data analysis technique was used in data mining. In the first step, in order to determine the number of clusters, additional calculations and normalization standardization method were applied to the dataset. Then, in terms of sum square error value, k was determined as equal 5: cluster 1 consists of 48 provinces, cluster 2 and cluster 3 has consist of 1 province, cluster 4 contains 11 provinces, and cluster 5 includes 20 provinces. On the basis of the statistics, it was concluded that the effects of all variables on the three clusters are significant. As a result, clustering of provinces in terms of APV, TCL, and TCV per capita shows that there are significant differences in obtained clusters.

References

Adanacıoğlu, H., Artukoğlu, M., & Güneş, E. (2017). Türkiye'de Tarımsal Kredi Performansının Çok Boyutlu Ölçekleme Yaklaşımıyla Analizi. *Turkish Journal of Agricultural Economics*, *23*(2).

Chandio, A. A., Jiang, Y., Gessesse, A. T., & Dunya, R. (2017). The nexus of agricultural credit, farm size and technical efficiency in Sindh, Pakistan: A stochastic production frontier approach. Journal of the Saudi Society of Agricultural Sciences.

Chen, Y. L., Chen, J. M., & Tung, C. W. (2006). A data mining approach for retail knowledge discovery with consideration of the effect of shelf-space adjacency on sales. *Decision support systems*, *42*(3), 1503-1520.

e Saqib, S., Ahmad, M. M., Panezai, S., & Ali, U. (2016). Factors influencing farmers' adoption of agricultural credit as a risk management strategy: The case of Pakistan. *International journal of disaster risk reduction*, *17*, 67-76.

Hayran, S., & Gül, A. (2018). Mersin İlinde Çiftçilerin Tarımsal Kredi Kullanım Kararlarını Etkileyen Faktörler. Iğdır Üniversitesi Fen Bilimleri Enstitüsü Dergisi, 8(1), 271-277.

Herbin, M., Bonnet, N., & Vautrot, P. (2001). Estimation of the number of clusters and influence zones. Pattern Recognition Letters, 22(14), 1557-1568.

Kamber, M., & Pei, J. (2006). Data Mining. Morgan kaufmann.

Kijewska, A., & Bluszcz, A. (2016). Research of varying levels of greenhouse gas emissions in European countries using the k-means method. Atmospheric Pollution Research, 7(5), 935-944.

Nisbet, R., Elder, J., & Miner, G. (2009). Handbook of statistical analysis and data mining applications. Academic Press.

Pham, D. T., Dimov, S. S., & Nguyen, C. D. (2005). Selection of K in K-means clustering. *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science*, *219*(1), 103-119.

Terin, M., Güler, İ. O., & Aksoy, A. (2014). Türkiye'de tarımsal üretim ile tarımsal kredi kullanımı arasındaki nedensellik ilişkisi. Iğdır Üniversitesi Fen Bilimleri Enstitüsü Dergisi, 4(1), 67-72.