SAKARYA ÜNİVERSİTESİ

FEN BİLİMLERİ ENSTİTÜSÜ DERGİSİ

Sakarya University Journal of Science
SAUJS

ISSN 1301-4048 | e-ISSN 2147-835X | Period Bimonthly | Founded: 1997 | Publisher Sakarya University | http://www.saujs.sakarya.edu.tr/

Title: Classification of Forest Fires in European Countries by Clustering Analysis Techniques

Authors: Hakan SERİN, Muslu Kazım KÖREZ, Mehmet Emin TEKİN, Sinan SİREN

Recieved: 2023-04-26 00:00:00

Accepted: 2023-06-21 00:00:00

Article Type: Research Article

Volume: 27
Issue: 5
Month: October
Year: 2023
Pages: 987-1001

How to cite
Hakan SERİN, Muslu Kazım KÖREZ, Mehmet Emin TEKİN, Sinan SİREN; (2023), Classification of Forest Fires in European Countries by Clustering Analysis Techniques. Sakarya University Journal of Science, 27(5), 987-1001, DOI: 10.16984/saufenbilder.1288073
Access link
https://dergipark.org.tr/tr/journal/1115/issue/80257/1288073

Sakarya University Journal of Science 27(5), 987-1001, 2023

Classification of Forest Fires in European Countries by Clustering Analysis Techniques

Hakan SERİN*[1], Muslu Kazım KÖREZ[2], Mehmet Emin TEKİN[1], Sinan SİREN[3]

Abstract

The biggest threat to the forests, which are natural habitats in European countries, as they are in the whole world, is forest fires. The aim of this study is to group the 38 European countries which have completely accessible fire indexes between the years 2008 to 2022; with respect to their similarities in fire regimes; and to compare the obtained groups with respect to their fire indexes. The clustering technique, which is a data mining method, was used while making these comparisons since it would be more objective and realistic to group and evaluate the countries according to their similarities. In the K-Means technique 2 clusters, and in the Ward's method 3 clusters were obtained. In the K-Means technique, significant statistical differences were found between the 2 clusters in terms of all fire indexes ($p<0.05$). In the Ward's method, statistically significant differences were found between the clusters in terms of the number of fires, total area burned (ha) and woodland ($p<0.05$). In the result of the studies, the fire regimes in Turkey, Bosnia and Herzegovina, Ukraine, Italy, Spain, and Portugal resulted higher than the other countries in both clustering algorithms. Since many factors were taken into consideration in the study, countries heavily associated with fires such as Greece and France were separated from those with high fire regimes. It is recommended to conduct modelling studies with data mining algorithms by taking different fire indexes into account in order to increase the reliability of the results.

Keywords: Data mining method, ward method, k-means, cluster analysis, forest fire

1. INTRODUCTION

Forests play an active role in protecting the balance of nature, contributing to the country's economy, increasing biodiversity, and combating global warming. Fires are the largest factor affecting forest presence in the world. When forest fires cannot be bring under control they can cause serious damage to bio-habitats, as well as vegetation; and cause the forest ecosystem to deteriorate [1, 2]. Although forest fires are a problem everywhere in the world, they may differ from region to region [3]. European countries have

* Corresponding author: hakan.461995@gmail.com (H. SERİN)
[1] Selcuk University, Faculty of Veterinary Medicine, Department of Biostatistics, Konya, Türkiye
[2] Selcuk University, Faculty of Medicine, Department of Biostatistics, Konya, Türkiye
[3] Selcuk University, Faculty of Veterinary Medicine, Department of Physiology, Konya, Türkiye
E-mail: mkkorez@gmail.com, mtekin@selcuk.edu.tr, sinan.siren1996@gmail.com
ORCID: https://orcid.org/0000-0002-1290-4547, https://orcid.org/0000-0001-9524-6115, https://orcid.org/0000-0002-3449-9984, https://orcid.org/0000-0003-2182-5047

developed forest management programs in order to protect the existence of forests covering 110 million hectares (ha), and the biodiversity in them. However, 500.000 hectares of land are affected by forest fires in Europe every year [4, 5]. Although forest fires that only destroy a small area are beneficial for the renewal of nature, the negative effects are much worse when the fires spread over larger areas. Global warming, which has increased in recent years, and the resulting climate changes are altering forest fire regimes. The increase in dry vegetation in forests due to increased drought can be given as an example of this situation. All these reasons force countries to develop effective fire management strategies in the fight against forest fires [6].

The fires in forest areas not only cause economic damage to countries but also destroy the benefits of nature to human health. Fires cause adverse reactions in some properties of the soil, both damaging the vegetation and also increasing the risks of natural disasters such as erosion and landslides. As a result of all these, the living spaces of the local people residing in the region are damaged, causing people to leave the region and thus change the social structure [7, 8].

Average amount of burning area and the average number of fires in the countries between 2010 and 2019 is given in Table 1. When Table 1 is examined, it is seen that Mediterranean countries are mostly in the top 5 among 26 European countries when the 10-year average amount of burned area and the number of fires is taken into account. The climate of the region, its vegetation, human-induced causes (arson, neglect), and tourism are conspicuous among the main reasons for this situation.

Table 1 Average amount of burning area and the average number of fires in the countries between 2010 and 2019 [9]

Ranking	**Country**	**Burning Annual Average Amount of Area (ha)**	**Ranking**	**Country**	**Annual Average Number of Fires**
1	Portugal	134307.6	1	Portugal	16800
2	Spain	94513.8	2	Spain	11859.9
3	Italy	63907.2	3	Poland	7188.2
4	Greece	24220.3	4	Italy	5419.6
5	France	12163.3	5	Sweden	4521.3
6	Croatia	11241.2	6	France	3907.4
7	Turkey	7332.1	7	Turkey	2477.3
8	Bulgaria	5266.4	8	Ukraine	1626.3
9	Hungary	4742.4	9	Czech Republic	1275.1
10	Sweden	4700	10	Finland	1260
11	North Macedonia	4473.5	11	Hungary	1218.3
12	Ukraine	3369.1	12	Greece	949.8
13	Poland	3027	13	Germany	864.7
14	Romania	1830.5	14	Latvia	580.5
15	Cyprus	1578.8	15	Bulgaria	470.8
16	Norway	1067.5	16	Romania	297.4
17	Germany	758.8	17	Slovakia	234.5

Table 1 Average amount of burning area and the average number of fires in the countries between 2010 and 2019 (Continue)

Ranking	**Country**	**Burning Annual Average Amount of Area (ha)**	**Ranking**	**Country**	**Annual Average Number of Fires**
18	Latvia	612.1	18	Austria	213.9
19	Finland	523.4	19	Norway	209.6
20	Slovakia	427.3	20	North Macedonia	209
21	Czech Republic	347.1	21	Croatia	199.4
22	Slovenia	270.5	22	Holland	181.8
23	Lithuania	117.5	23	Lithuania	152
24	Holland	112.1	24	Cyprus	102.7
25	Switzerland	107.6	25	Switzerland	98.4
26	Austria	66.5	26	Slovenia	83.1

There are 2 main factors in the emergence of forest fires in the world. These are lightning strikes (1%) and human-caused factors (99%). Human-caused fires occur intentionally or because of negligence. More than 95% of fires in Europe are a resulted from human-caused factors [5, 10]. The climate and vegetation in the region play a decisive role in the damage caused by forest fires. In regions with Mediterranean-type climates drought and high temperatures cause fires to be more common [11, 12].

Factors such as the migration of people from rural settlements and the limited use of low-yielding forest areas in Mediterranean coastal countries cause an increase in the density of highly flammable vegetation in forests. Therefore, when all these factors are combined with the heat waves in the summer months and human- induced causes, even an ineffectual fire factor can cause the burning of vast areas [13, 14]. It is expected that climate change will significantly increase the number of fires and the amount of burned forest areas; because of increasing drought, heat waves, and strong winds; especially in European countries with a coast to the Mediterranean Sea. In addition, the decrease in the rural population and the change in the demographics of the regions with the increase in tourism-oriented compositions are also important factors in the incrementation of forest fires [15, 16]. The number of fires in Turkey in 2021 according to their causes are given in Figure 1. In Figure 1, when we look at the number of fires that occurred in 2021 in Turkey, with respect to their reason for starting; it is seen that approximately 40% of the fires are caused by humans. However, this rate can be much higher when fires with unknown causes are considered.

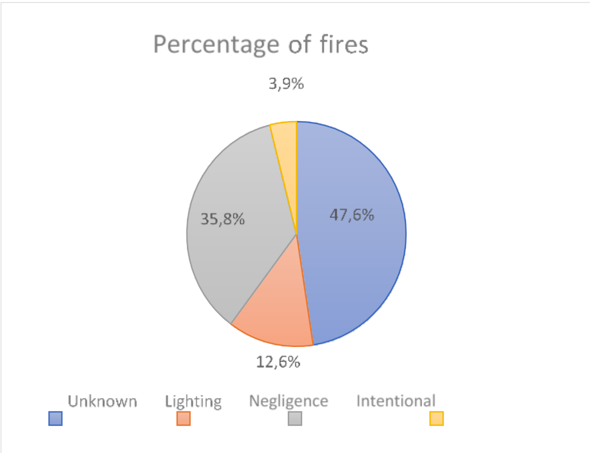

Figure 1 Number of fires in Turkey in 2021 according to their causes [17]

The classification or clustering of forest fires plays an important role for countries to take a measure against fire. There are many studies that use of clustering and classification algortihms to predict and discriminate of forest fires. Junior et al. clasifiy the fire risk situation in summer and winter months with meteorological fire indicators of 310 regions in the European Nomenclature of territorial units for statistics (NUTS) 3 region with K-means, Fuzzy K-means and Gaussian mixture models. More accurate predictions are obtained with Gaussian mixture models in summer, and K-means method in winter [18]. Cortez and Morais estimate the amount of burnt area using Multiple Regression (MR), Decision Trees (DT), Neural Networks (NN), Support Vector Machines (SVM) and Random Forests (RF) algorithms in spatial and temporal fire indicators show that the algorithms are predicted small fires more accurately than large fires and the best performance is obtained with the SVM method [19].

Li et al. develop an algorithm for automatic identification of fire smoke with artificial neural networks using high resolution radiometer images. Thus, they determine that possible forest fires can be put out much faster before the severity of the fire increases [20]. Han et al. estimate the forest fire hazardous area with the likelihood ratio and conditional probability methods. According to the results of FHR (Forest Fire Hazard Rate) and PRC (Prediction Rate Curve), more accurate predictions are obtained with the likelihood ratio method than the conditional probability method [21]. Shidik and Mustofa used Fuzzy C-Means and Back-Propagation Neural Network methods to classify burned forest areas and obtained 97.5% correct classification rate [22].

Rosadi and Andriyani compare the performances of Decision Trees (DT), Support Vector Machines (SVM), Fuzzy c-means and Adaptive Boosting (AdaBoost) algorithms using the variables of burned forest area, number of fires and their ratio to area. The most accurate classification performance is obtained from AdaBoost and the lowest is obtained from SVM algorithm [23]. Sinha et al. use a semi-supervised rule-based classification model to classify fires according to their severity and obtain a model accuracy of 96% [24]. Tutmez et al. cluster the number of forest fires and the amount of burnt area in Antalya using the K-Mod clustering method [25]. Yin et al. classify the fire incidence in forest, grassland, cropland and bare land areas in the Central Asia region using the K-Means method [26].

In this study, it is aims to group European countries according to their fire regimes using data mining methods and to compare the obtained groups by considering fire factors. The study is a preliminary information for the researchers who search the literature on forest fire. In addition, it is an example for the classification of forest fires with machine learning and a guide for future studies. When the literature is analysed, the performances of machine learning algorithms on forest fires are generally compared. However, it is seen that the contribution of the obtained results to the application area is limited. In this study, not only the algorithm classification but also the intersection and differentiation aspects of the classifications made by these algorithms in an applied field with the literature are discussed and a different perspective is brought to the field.

2. MATERIALS AND METHODS

2.1. Study Area

In the study, data covering the years 2008-2022 are used. In order to group European countries according to the degree of exposure to forest fires, clustering analysis is performed by using the number of fires, total area burned (ha), woodland (km^2) and burnt area by fire.

Table 2 Fire indicators and units used in the study

Fire Indexes	Unit
Number of fires	Number
Total area burned	Hectare
Woodland	km^2
Burnt area by fire	Hectare

While selecting the variables used in the study, variables directly related to forest fire are selected in order to make an objective and accurate classification (Table 2). In addition, machine learning studies on forest fires are analysed and taken into consideration for variable selection [27, 28]. Data are obtained from the "Current Statistics Portal" of the EFFIS (European Forest Fire Information System) database [29]. The sample of the study consists of 38 European countries, 27 of which are members of the European Union and 11 of which are not. These are Austria, Belgium, Bulgaria, Croatia, Cyprus, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Ireland, Italy, Latvia, Lithuania, Luxembourg, Malta, Netherlands, Poland, Portugal, Romania, Slovakia, Slovenia, Spain, and Sweden which are members of the European Union. The countries that are not members of the European Union are Albania, Andorra, Bosnia and Herzegovina, Macedonia, Montenegro, Norway, Serbia, Switzerland, Turkey, Ukraine, and the United Kingdom.

2.2. Statistical Analysis

All statistical analyses are performed with R statistical language (version 4.2.1; The R Foundation for Statistical Computing, Vienna, Austria; https://www.r-project.org). K-Means technique and Ward's method for the hierarchical clustering method are used in the study. The Elbow method is used to determine the optimum number of clusters in K-Means technique. In this study, the most widely used Euclidean distance is used to determine the distance matrix [30]. In the clustering technique, data on forest fire indexes are standardized and included in the analysis.

In order to apply the K-Means technique used in the study, the data should be of continuous type. In this method, units are classified by taking into account their distance to the cluster mean. K-Means technique gives consistent results in large data masses. However, the disadvantage of this method is that it is sensitive to outliers and the number of clusters is determined by the researcher. In Ward technique, the other technique used in the study, the number of clusters is decided by looking at the dendrogram graph. Ward technique is more statistical than other techniques since it performs variance-based merging. The disadvantage of this technique is that if the researcher is inexperienced, he may misinterpret the dendrogram graph and determine the number of clusters differently.

Shapiro–Wilk test and Q-Q plots graphs are used to assess the normality of the data in the groups obtained in the result of the cluster analysis. Mann–Whitney U test and Kruskal-Wallis test followed by Dunn's post-hoc test with Bonferroni correction are run to determine whether there is a statistically significant difference in the number of fires, total area burned, woodland and burnt area by fire between clustering. A *p*-value less than 0.05 is considered statistically significant.

2.3. Clustering Methods

Data mining is an important process that allows the discovery of hidden knowledge from large amounts of data. Clustering analysis is an important data mining method that allows dividing a data set with a mixed structure into homogeneous subgroups by reducing the size. In today's technology, clustering analysis is often used in databases as an unsupervised learning method for the analysis of large sized data [30, 31]. There is no predetermined classification process in this analysis method. The aim of clustering techniques is to separate homogeneous clusters within clusters and heterogeneous clusters between clusters.

When determining the number of clusters in clustering methods, a number of clusters can be suggested by the researcher by considering the theoretical structure in the K-Means technique or methods for determining the optimum number of clusters can be used. In hierarchical clustering analysis, on the other hand, the distance matrix between the objects is calculated and the objects that are similar are combined to form clusters, the number of clusters is not predetermined [31, 32]. The sum of the squared values of distance from the centre of the cluster to which each data is assigned according to its distance from the cluster centre is taken, and obtaining a cluster composition in which the sum of squares within the cluster are minimum is attempted [30, 33]. Clustering techniques are depend on creating homogeneous groups using distance matrices. Distance matrices such as Euclidean, Manhattan, and Minkowski are used for the distance between observations in the clustering process. The most widely used of these is the Euclidean distance. The formula for calculating the Euclidean distance is given in the equation below [34].

$$d_{Euclidian}(x,y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} \qquad (1)$$

Ward's method is the most commonly used hierarchical technique. This method differs from other techniques as clusters are joined based on variances. It is a method based on merging clusters whose sum of error squares is minimum [31, 32].

In hierarchical clustering methods; the number of clusters can be decided by examining the dendrogram graph of the distance between clusters, or it can be determined as the optimum number of clusters by examining the breakpoint of the graph obtained by dividing the in-group variance by the intergroup variance. In addition, another method proposed by Lewis and Thomas is to determine it according to the total variance and according to the contribution of the added set to the variance [30]. The determined clusters should have the ability to explain 80% of the variance and the added cluster should have a contribution of at least 5% to the total variance [30, 35]. However, all clustering methods, there is no definite criterion for determining the number of clusters, so it can be undecided. Therefore, some formulas have been developed for calculating the number of clusters. The most used one is "$k = \sqrt{n/2}$" [36, 37].

3. RESULTS AND DISCUSSION

3.1. Results of K-Means Cluster Analysis with Fire Indexes of European Countries

In the study, fire indexes are examined with the K- Means clustering technique. In order to determine the optimal number of clusters; information about the countries such as their location, climate, vegetation coverage, and fire statistics are taken into account. In addition, the number of clusters is decided by using the methods of determining the optimal number of clusters. The results are obtained in the R statistical program; and "dplyr", "stats", "cluster", "clusterR", and "ggplot2" packages are used.

The optimal number of clusters is determined by using mathematical measurement methods (Elbow and Silhouettes techniques) [34, 38]. In Figure 2, the results of the Elbow and Silhouettes techniques are given.

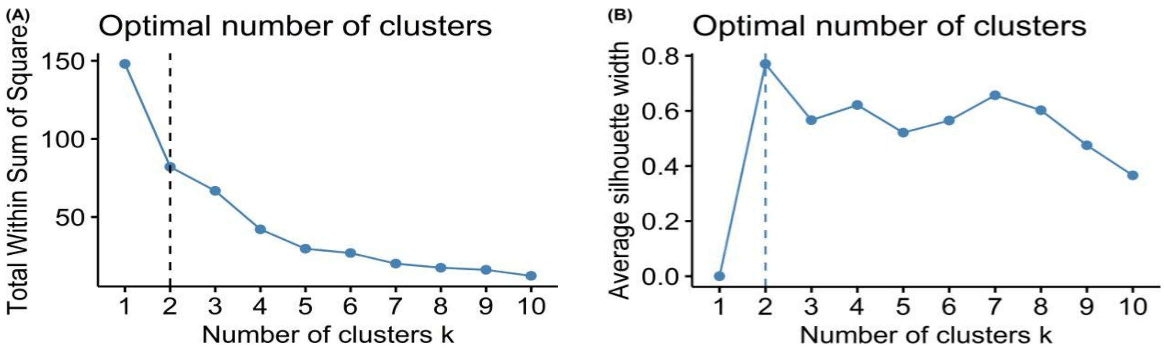

Figure 2 Determination of the optimal number of clusters using the Elbow (A) and Silhouette (B) methods

The number of clusters is determined as 2 since the total value of the squares within the cluster in the elbow graph gradually decreases after two and considering the peak of the silhouette width in the silhouette graph. The number of observations in the clusters is 32 and 6, respectively, and the cluster compositions are shown in Figure 3. In Figure 3, after the clustering analysis, the cluster centres and the findings of the two clusters are given in two dimensions. According to these results, it is seen that the majority of European countries, including all Northern European countries, are in the first cluster. The countries of Turkey, Ukraine, Italy, Spain, Portugal, and Bosnia and Herzegovina are in the second cluster. The results comparing the fire indexes of the two clusters are given in Table 3.

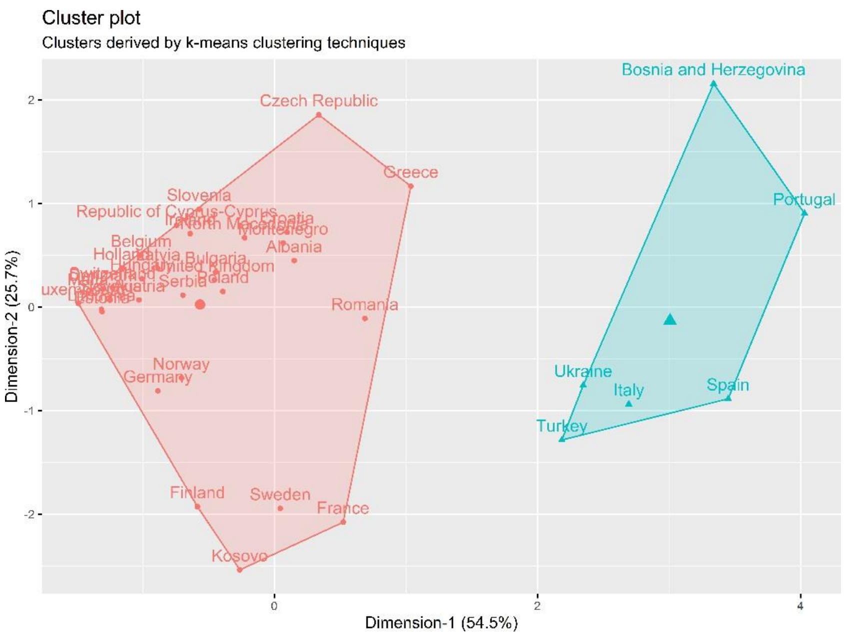

Figure 3 Clusters obtained by fitting K-Means clustering based on fire indexes

Table 3 Comparison of clusters by fire indexes

Fire Indexes	Cluster 1 (*n*=32)	Cluster 2 (*n*=6)	*p*-value
Number of fires	75 (9.25 – 545)	3149.5 (1893.50 – 3675.50)	<.001
Total area burned (Ha)	11436 (1365.75 – 113737.25)	981322.5 (746628.25 – 1333282.50)	<.001
Woodland (km^2)	23983 (7584 – 62107.5)	105868 (30264.75 – 192330.25)	.045
Burnt area by fire (Ha)	195.5 (122.5 – 290.75)	383.5 (203.50 – 576.25)	.029

p-values are calculated by Mann-Whitney U test.
Data are presented as median with interquartile range (25th percentile – 75th percentile).
p< .05 is considered statistically significant.

As a result of the Mann-Whitney U test performed on fire indexes which are divided into two clusters, a significant difference is found between the two clusters in terms of all indexes (Figure 4). The results obtained in Table 2 support that the two cluster compositions are completely divided from each other. When the number of fires, total area burned, woodland and burnt area by fire in Turkey, Ukraine, Italy, Spain, Portugal, and Bosnia and Herzegovina countries in Cluster 2 are examined, it is found that these numbers are significantly higher than other European countries.

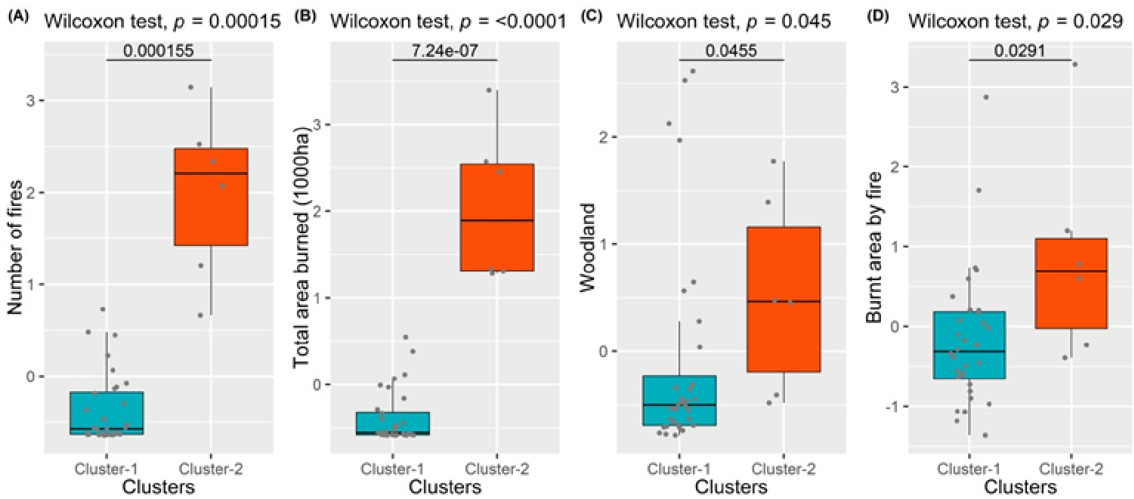

Figure 4 Comparison of the clusters in terms of the number of fires (A), total area burned (B), woodland (C) and burnt area by fire (D). Data are presented as standardized values.

3.2. Results of Hierarchical Cluster Analysis with Forest Fire Indexes of European Countries

In addition to the K-Means clustering technique, Hierarchical clustering techniques are also applied to the forest fire indexes are discussed in the study, and the results are evaluated. In hierarchical clustering techniques, the number of clusters is decided by evaluating the dendrogram graph [31]. Cluster compositions are obtained through the R program language, and the packages "dendextend", "purrr", "gridExtra", "ggpubr", "factoextra", "hclust", "dplyr", and "tidyverse" are used.

In this study, hierarchical clustering techniques called linkage and variance techniques are used. Among the linkage methods, "single connection", "average connection", and "complete connection" techniques are used; while Ward's method are used in the variance technique. In order to choose the most fitting method, the agglomeration coefficient of each technique is examined and obtained as 0.76 for single connection, 0.82 for average connection, 0.88 for full connection, and 0.93 for the Ward's method. Since the closer the agglomeration coefficient is to 1 the stronger the cluster composition will be, the analysis is continued using the Ward's method. The dendrogram graph obtained by Ward’s method is given Figure 5. Ward's method is preferred also because it gives more accurate results in studies that try to minimize in-cluster variability, and where there are small sample size [30, 37].

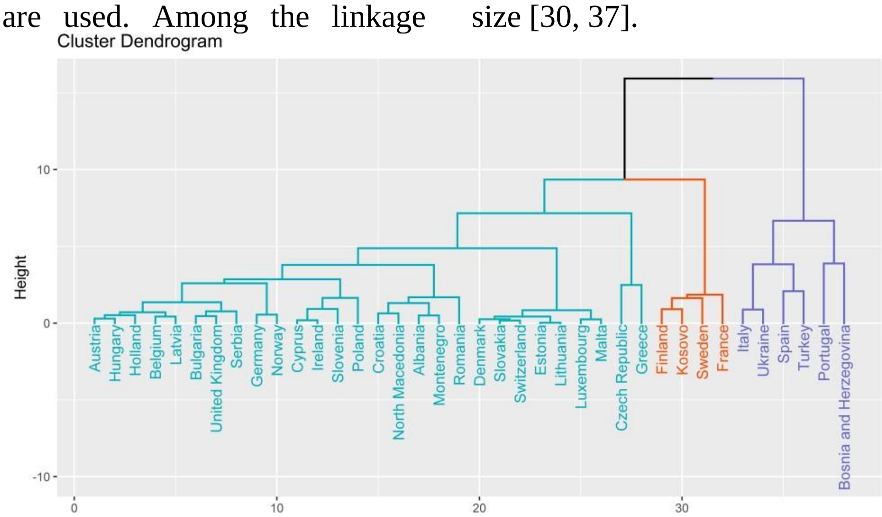

Figure 5 Ward dendrogram of fire indexes

When Figure 5 is examined, it shows that it is ideal to divide the countries included in the study into 3 clusters. When these clusters are examined, it is seen that the first cluster consists of 28 countries, the second cluster consists of 4 countries, and the third cluster consists of 6 countries. The clusters are compared with the Kruskal-Wallis H test in order to better analyze the forest fire indexes of the determined clusters and to examine the heterogeneity between clusters with other statistical methods. The results are given in Table 4 and Figure 6.

Table 4 Comparison of clusters by fire indexes

Fire Indexes	Cluster 1 (*n*=28)	Cluster 2 (*n*=4)	Cluster 3 (*n*=6)	*p*-value*
Number of fires	44.5 (7.25 – 545)^{b}	214 (48 – 980)^{ab}	3149.5 (1893.5 – 3675.5)^{a}	<.001
Total area burned (Ha)	8967.5 (1257.25 – 113737.25)^{b}	41074.5 (12666.75 – 174710.25)^{ab}	981322.5 (746628.25 - 1333282.5)^{a}	<.001
Woodland (km^2)	21223 (7540 – 36250)^{b}	263685 (236650 – 286182.5)^{a}	105868 (30264.75 – 192330.25)^{a}	<.001
Burnt area by fire (Ha)	205.5 (106.75 – 290.75)^{a}	164.5 (138.75 – 289.25)^{a}	383.5 (203.5 – 576.25)^{a}	.085

p-values are calculated by Kruskal-Wallis test.
Different small superscript letters in each row denote that statistically significant difference in post-hoc analysis
Data are presented as median with interquartile range (25th percentile – 75th percentile).
p< .05 is considered statistically significant.

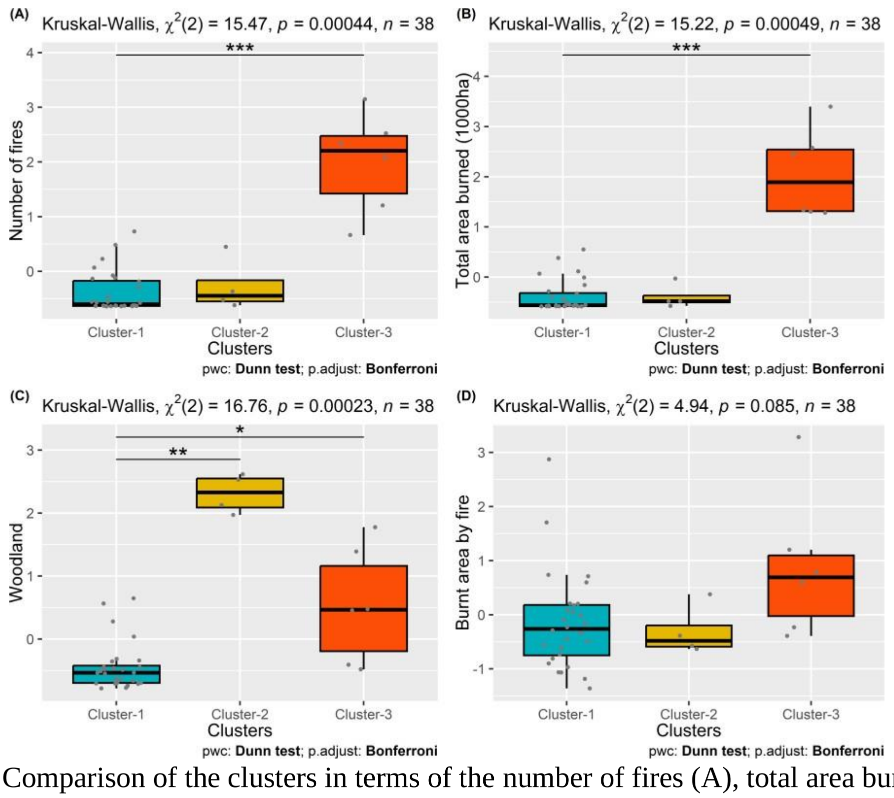

Figure 6 Comparison of the clusters in terms of the number of fires (A), total area burned (B), woodland (C) and burnt area by fire (D). Data are presented as standardized values.

Considering to the comparison of clusters according to the fire indexes using Kruskal-Wallis test, there is a statistically significant difference in the number of fires ($\chi2$=15.47, *p*=<.001), total area burned ($\chi2$=15.47, *p*=<.001) and woodland ($\chi2$=15.47, *p*=<.001) between clusters. Dunn post-hoc test reveal that the number of fires (3149.5 [IQR, 1893.5 – 3675.5] vs. 44.5 [IQR, 7.25 – 545], adj. *p*-value<.001) and total area burned (981322.5 [IQR, 746628.25 – 1333282.5] vs. 8967.5 [IQR, 1257.25 – 113737.25], adj. *p*-value<.001) in Cluster-3 countries are significantly higher than that in Cluster-1 countries. In addition to, Woodland in Cluster- 2 (263685 [IQR, 236650 –

286182.5]) countries are significantly larger compared to Cluster-1 (21223 [IQR, 7540 – 36250], adj. *p*-value<.001) and Cluster-3 (105868 [IQR, 30264.75 – 192330.25], adj. *p*- value=.012) countries (Figure 6). When the two clusters that formed in the result of the K-Means cluster analysis are examined, Turkey, Bosnia and Herzegovina, Ukraine, Italy, Spain, and Portugal are in the same cluster; while the remaining European countries form the other cluster. As a result of the hierarchical clustering analysis performed with the Ward's method, the countries are divided into 3 clusters. Finland, Kosovo, Sweden, and France formed a cluster; and Turkey, Bosnia and Herzegovina, Ukraine, Italy, Spain, and Portugal formed a separate cluster. The remaining European countries constitute the third cluster.

In various studies on forest fires in European countries Various in studies on forest fires in European Countries [5, 11, 14-16, 39, 40] it is reported that forest fires are concentrated in France, Portugal, Spain, Greece, and Italy in the Mediterranean region. When the results are interpreted in the light of the literature, it is seen that this situation is caused by reasons such as climate, vegetation, arson, neglect, and tourism in Mediterranean countries (Turkey, Italy, Spain, Portugal, France) [5, 14]. Bosnia and Herzegovina has a climate structure with different climate types, including the Mediterranean climate. In addition, seasonal factors such as precipitation regime and temperature increase are shown to play an important triggering role in fires [41]. In the two clusters structure preferred in the K-Means clustering method in this study, Spain, Portugal and Italy are in the first cluster, while Greece and France are in the second cluster, although they differ from other European countries. In the cluster composition obtained by the Ward's method; Spain, Portugal, and Italy are in the same cluster; France is clustered together with Finland, Kosovo, and Sweden; and Greece is in the same cluster with the other European countries. Therefore, although there are some differences, it is seen that the results are compatible with the literature. When the literature is examined [9, 42], it is seen that the fire regime in countries such as Finland, Sweden, Turkey, and Ukraine differs from other European countries. These studies support the findings obtained as a result of cluster analysis. In both the K-Means and Ward's methods, the country of Bosnia and Herzegovina is in the same cluster with countries such as Spain, Portugal, and Italy; while countries of Finland and Kosovo are in the same group with France in the Ward's method. In addition, the most surprising result of the study is that Greece is included in the group dominated by low-fire regimes, which includes the majority of European countries in both clustering techniques (Figure 3, Figure 5). The general thought is that this difference occurs due to the many factors that are examined together in clustering analysis.

4. CONCLUSIONS

Forest fires pose a serious problem for European countries as they do all over the world [5]. In this study, the similarities and differences between European countries in terms of forest fire indexes are examined in light of it being an ongoing and interesting subject. In the study, clustering analysis, which is a data mining method, was carried out to segment 38 European countries, whose data are fully accessible, according to their forest fire indexes. In the cluster analysis carried out for this purpose, the clusters formed by the countries are examined and the countries are classified according to their fire regimes in terms of the factors included in the analysis. Then, the clusters are compared according to their fire indexes and the results are evaluated.

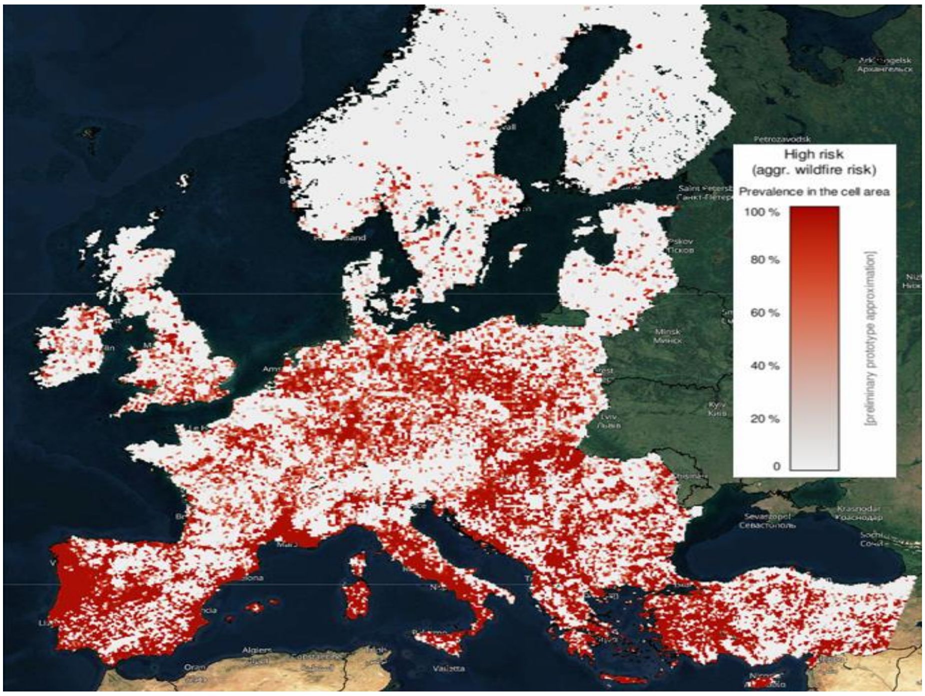

Figure 7 EFFIS Forest Fire Risk Map for European countries [18]

In this study, unlike similar studies in the literature, the results obtained from the classification algorithms are tried to be interpreted both in terms of theoretical results and in the field of forest fires. In this respect, the study constitutes a good example of the application of machine learning results to an applied science. It also serves as a preliminary information for future research in this field and guides the researcher in the process of research design. Since the study compares different clustering techniques, it gives the researcher the opportunity to interpret the classification principle of these techniques.

The study also gives the researcher an idea about how effective European countries are in fire planning. Countries in the same region should take into account the fire planning of countries that are in the same risk group but suffer less damage from forest fires when developing fire management strategies. For example, although Italy and Croatia have similar vegetation cover, Italy is in Cluster-3 with high fire regime while Croatia is in Cluster-1 with low fire regime. Therefore, Italy's fire management plans and penal sanctions are insufficient to prevent forest fires.

When Figure 7 is examined, the findings obtained as a result of the study coincide with the European fire map by EFFIS. The study is carried out by taking into account only the countries included in the EFFIS database; and the countries whose data cannot be accessed are excluded from the scope of the study. Unexpected results may occur, since cluster analysis makes an objective evaluation by considering many factors rather than a subjective point of view. The scope of the study can be expanded by increasing the number of countries and the fire indexes used, and different clustering methods can be used for estimation. In addition, it is foreseen that modeling studies can be carried out using data mining algorithms and estimation variables.

Funding
The author (s) has no received any financial support for the research, authorship or publication of this study.

Authors' Contribution
The authors contributed equally to the study.

The Declaration of Conflict of Interest/ Common Interest

No conflict of interest or common interest has been declared by the authors

The Declaration of Ethics Committee Approval

This study does not require ethics committee permission or any special permission.

The Declaration of Research and Publication Ethics

The authors of the paper declare that they comply with the scientific, ethical and quotation rules of SAUJS in all processes of the paper and that they do not make any falsification on the data collected. In addition, they declare that Sakarya University Journal of Science and its editorial board have no responsibility for any ethical violations that may be encountered, and that this study has not been evaluated in any academic publication environment other than Sakarya University Journal of Science.

REFERENCES

[1] E. A. Johnson, K. Miyanishi, "Forest fires: behavior and ecological effects". Elsevier, Amsterdam, 2001.

[2] E. Gutiérrez, S. Lozano, "Avoidable damage assessment of forest fires in European countries: an efficient frontier approach," European journal of forest research, vol. 132, no. 1, pp. 9-21, 2013.

[3] J. G. Pausas, "Changes in fire and climate in the eastern Iberian Peninsula (Mediterranean basin)", Journal of Climatic change, vol. 63, no. 3, pp. 337-350, 2004.

[4] F. Europe, F. Unece, "State of Europe’s forests 2015," presented at the Ministerial Conference on the Protection of Forests in Europe, 04.11.2015, 2015.

[5] N. Khabarov, A. Krasovskii, M. Obersteiner, R. Swart, "Forest fires and adaptation options in Europe," Journal of Regional Environmental Change, vol. 16, no. 1, pp. 21-30, 2016.

[6] M. Grünig, R. Seidl, C. Senf, "Increasing aridity causes larger and more severe forest fires across Europe," Journal of Global Change Biology, Dec 14 2022.

[7] A. Hamit, K. Ali, C. Huseyin, "The structure and importance of fire suppressing organizations in Turkey," Journal of Scientific Research and Essays, vol. 5, no. 5, pp. 456-460, 2010.

[8] Z. G. Yılmaz, "Yaşar Kemal ve Orman Yangınları," Ekoeleştiri Çevre ve Edebiyat. Phoenix Yayınevi, Ankara, pp. 129-169, 2012.

[9] C. Şahan, K. İsmail, "Türkiye ve Avrupa Ülkeleri Orman Yangınlarının Bazı Değişkenler Açısından Değerlendirilmesi," Ordu Üniversitesi Sosyal Bilimler Enstitüsü Sosyal Bilimler Araştırmaları Dergisi, vol. 12, no. 1, pp. 1-14, 2022.

[10] A. Küçükosmanoğlu, "Ülkemizde orman yangınları ve yangın sezonları," Journal of the Faculty of Forestry Istanbul University, vol. 44, no. 1-2, pp. 121-128, 1994.

[11] A. Dimitrakopoulos, C. Gogi, G. Stamatelos, I. Mitsopoulos, "Statistical Analysis of the Fire Environment of Large Forest Fires (> 1000 ha) in Greece," Polish Journal of Environmental Studies, vol. 20, pp. 327-332, 01/01 2011.

[12] O. Kahraman, O. Esin, G. Çamalan, "İzmir-Tırazlı orman yangınının uydu ve model verileri ile analizi," Ulusal Çevre Bilimleri Araştırma Dergisi, vol. 4, no. 1, pp. 1-12, 2021.

[13] F. Moreira, O. Viedma, M. Arianoutsou, T. Curt, N. Koutsias, "Landscape–wildfire interactions in southern Europe: implications for landscape management," Journal of environmental management, vol. 92, no. 10, pp. 2389-2402, 2011.

[14] J. San Miguel Ayanz, J. M. Moreno, A. Camia, "Analysis of large fires in European Mediterranean landscapes: Lessons learned and perspectives," Journal of Forest Ecology and Management, vol. 294, pp. 11-22, 2013.

[15] J. San Miguel Ayanz, M. Rodrigues, S. Oliveira, C. Pacheco, F. Moreira, "Land cover change and fire regime in the European Mediterranean region," in Journal of Post-fire management and restoration of southern European forests: Springer, 2012, pp. 21-43.

[16] F. Rego, E. Rigolot, P. Fernandes, C. Montiel, J. Silva, "Towards integrated fire management. European Forest Institute Policy Brief 4," Joensuu, Finland), 2010.

[17] T. C. T. v. O. B. O. G. M. "Ormancılık İstatistikleri "(accessed 26.12.2022, 2022).

[18] J. S. Júnior, J. R. Paulo, J. Mendes, D. Alves, L. M. Ribeiro, C. Viegas, "Automatic forest fire danger rating calibration: Exploring clustering techniques for regionally customizable fire danger classification," Expert Systems with Applications, vol. 193, p. 116380, 2022.

[19] P. Cortez, A. d. J. R. Morais, "A data mining approach to predict forest fires using meteorological data," Department of Information Systems R&D Algoritmi Centre, University of Minho, 2007.

[20] Z. Li, Y. Kaufman, C. Ichoku, R. Fraser, A. Trishchenko, L. Giglio, J. Jin, X. Yu, "A review of AVHRR-based active fire detection algorithms: Principles, limitations, and recommendations," Global and regional vegetation fire monitoring from space, planning and coordinated international effort, pp. 199-225, 2001.

[21] J. G. Han, K. H. Ryu, K. H. Chi, Y. K. Yeon, "Statistics based predictive geo-spatial data mining: Forest fire hazardous area mapping application," in Web Technologies and Applications: 5th Asia-Pacific Web Conference, APWeb 2003, Xian, China, April 23–25, 2003 Proceedings 5, 2003: Springer, pp. 370-381.

[22] G. F. Shidik, K. Mustofa, "Predicting size of forest fire using hybrid model," in Information and Communication Technology: Second IFIP TC5/8 International Conference, ICT-EurAsia 2014, Bali, Indonesia, April 14-17, 2014. Proceedings 2, 2014: Springer, pp. 316-327.

[23] D. Rosadi, W. Andriyani, "Prediction of forest fire using ensemble method," in Journal of Physics: Conference Series, 2021, vol. 1918, no. 4: IOP Publishing, p. 042043.

[24] D. Sinha, R. Kumari, S. Tripathi, "Semisupervised classification based clustering approach in WSN for forest fire detection," Wireless Personal Communications, vol. 109, pp. 2561-2605, 2019.

[25] B. Tutmez, M. G. Ozdogan, A. Boran, "Mapping forest fires by nonparametric clustering analysis," Journal of forestry research, vol. 29, pp. 177-185, 2018.

[26] H. Yin, J. Guli, L. Jiang, T. Yu, J. Umuhoza, X. Li, "Monitoring fire regimes and assessing their driving

factors in Central Asia," Journal of Arid Land, vol. 13, pp. 500-515, 2021.

[27] J. Li, Y. Shan, S. Yin, M. Wang, L. Sun, D. Wang, "Nonparametric multivariate analysis of variance for affecting factors on the extent of forest fire damage in Jilin Province, China," Journal of Forestry Research, vol. 30, pp. 2185-2197, 2019.

[28] L. Iliadis, M. Vangeloudh, S. Spartalis, "An intelligent system employing an enhanced fuzzy c-means clustering model: Application in the case of forest fires," Computers and Electronics in Agriculture, vol. 70, no. 2, pp. 276-284, 2010.

[29] E. F. F. I. S. EFFIS. "EFFIS Statistics Portal." (accessed 10.11.2022, 2022).

[30] Y. K. Akın, "Veri madenciliğinde kümeleme algoritmaları ve kümeleme analizi," Marmara Üniversitesi / Sosyal Bilimler Enstitüsü / Ekonometri Ana Bilim Dalı, Marmara Üniversitesi / Sosyal Bilimler Enstitüsü / Ekonometri Ana Bilim Dalı, Turkey, 2008.

[31] B. Altunkaynak, "Veri madenciliği yöntemleri ve R uygulamaları". Seçkin Yayıncılık, Ankara, 2019.

[32] F. Murtagh, P. Legendre, "Ward’s hierarchical agglomerative clustering method: which algorithms implement Ward’s criterion?," Journal of classification, vol. 31, no. 3, pp. 274-295, 2014.

[33] H. Ehsan, D. A. Hamed, S. Jamal, "Application of data mining techniques in stock markets: A survey," Journal of Economics and International Finance, vol. 2, no. 7, pp. 109-118, 2010.

[34] A. Kassambara, "Practical guide to cluster analysis in R: Unsupervised machine learning". STHDA, 2017.

[35] F. J. M. Ruíz, "Strategic group analysis in strategic marketing: an application to Spanish savings banks", Marketing Intelligence & Planning, vol. 6, pp. 257-270, 1998.

[36] C. Alpar, Uygulamalı çok değişkenli istatistiksel yöntemler. Detay Yayıncılık, Ankara, 2017.

[37] C. Sungur, "Sağlık göstergelerine göre Ekonomik Kalkınma ve İşbirliği Örgütü ülkelerinin kümeleme analizi," SGD-Sosyal Güvenlik Dergisi, vol. 6, no. 1, pp. 197-224, 2016.

[38] Ö. F. Rençber, "Veri madenciliğinde kullanılan kümeleme algoritmalarının karşılaştırılması üzerine bir inceleme," Düzce Üniversitesi Bilim ve Teknoloji Dergisi, vol. 7, no. 3, pp. 1671-1685, 2019.

[39] P. Barbosa, A. Camia, J. Kucera, G. Liberta, I. Palumbo , "Assessment of forest fire impacts and emissions in the European Union based on the European Forest Fire Information System," Journal of Developments in Environmental Science, vol. 8, pp. 197-208, 2008.

[40] P. Mateus, P. M. Fernandes, "Forest fires in Portugal: dynamics, causes and policies," Journal of Forest context and policies in Portugal: Springer, 2014, pp. 97-115.

[41] B. Drašković, B. Mıletıć, M. Gutalj, S. Stjepanovıć, "Climate changes and fires in Bosnia and Herzegovina," in XI International Scientific Agriculture Symposium ,AGROSYM, Republic of Srpska, Bosnia and Herzegovina, October 2020, pp. 694-702.

[42] F. Tedim, G. Xanthopoulos, V. Leone, "Forest fires in Europe: Facts and

challenges," in Wildfire hazards, risks and disasters: Elsevier, 2015, pp. 77-99.