**VERİ MADENCİLİĞİ İLE ÜLKELERİN REFAH DÜZEYLERİNE GÖRE ANALİZİ**

***Öz***

*Veri Madenciliği (DM) büyük veri yığınlarında saklı olan anlamlı bilgileri ortaya çıkarmayı sağlayan teknikleri içermektedir. Uygulama odaklı bir alan olarak VM, istatistik, makine öğrenimi, örüntü tanıma, veritabanı ve veri ambarı sistemleri, bilgi alma, görselleştirme, algoritmalar, yüksek performanslı hesaplama ve birçok uygulama alanı gibi diğer alanlardan birçok tekniği birleştirmektedir. Bu tekniklerden biri olan Kümeleme Analizi (KA), doğal gruplamaları kesin olarak bilinmeyen, ham veri matrisinde bulunan birimleri sahip oldukları temel özellikler bakımından birbirleri ile benzer olan alt gruplara (kümelere) ayırmaktadır. Bunun için de hiyerarşik olan ve hiyerarşik olmayan olmak üzere iki grupta toplanan algoritmalar kullanılmaktadır. Bu çalışmada amaç, ülkeleri VM tekniklerinden biri olan KA ile refah düzeylerine göre gruplara ayırmak ve yöntemin uygulanabilirliğini göstermektir. Bu kapsamda, Legatum Refah Endeksi (LRE)’nde yer alan 12 refah göstergesi dahilinde 167 ülkenin 2019 yılı verilerinden yararlanılmıştır. Çalışmada, hiyeraşik KA yöntemlerinden Wards algoritmasıyla ülkeler gruplara ayrılmış, hiyeraşik olmayan KA yöntemlerinden K-Ortalamalar ile ülkeler arasındaki benzerlikler saptanmış ve Türkiye’nin gruplar içerisindeki yeri belirlenmiştir. Bulgular; ülkelerin refah düzeylerine göre üç kümeye ayrıldığını, kümelere ayırmada en etkili göstergelerin “pazar erişimi ve altyapı, eğitim, yatırım ortamı”, en az etkili göstergelerin “sosyal sermaye, doğal çevre, emniyet ve güvenlik” olduğunu göstermektedir. Kümeler bazında; en yüksek refah düzeyine sahip ülkelerin yer aldığı kümenin “yönetim, girişim koşulları, yatırım ortamı” göstergeleri yönüyle diğerlerinden daha güçlü, en düşük refah düzeyine sahip ülkelerin yer aldığı kümenin ise “eğitim, sağlık, yaşam koşulları” göstergeleri yönüyle diğerlerinden oldukça zayıf olduğu saptanmıştır. Türkiye’nin yer aldığı orta refah seviyesindeki kümede; “sağlık, yaşam koşulları, eğitim” göstergelerinin en yüksek, “doğal çevre, kişisel özgürlükler, yönetim” göstergelerinin ise en düşük olduğu belirlenmiştir. Sonuçlar, KA’nın ülkeleri refah seviyesine göre gruplara ayırmada etkin sonuçlar gösterdiğini ortaya koymaktadır. Buna göre; refah seviyesi en yüksek olan ülkelerde bireyler ve kurumlar arasındaki ilişkinin ve uyumun güçlü olduğu, yenilik, yatırım, iş ve ticaret teşvikleriyle büyümenin kolaylaştığı; refah seviyesi en düşük olan ülkelerde sosyal refahı temsil eden, bireylerin özerklik ve tam potansiyellerine ulaşmalarını sağlayan yönlerin oldukça zayıf olduğu söylenebilir. Türkiye’nin sosyal refahı temsil eden konularda orta düzeyde, bireylerin temel özgürlükleri ve kendini geliştirebilme konularında ise zayıf olduğu görülmüştür.*

***Anahtar kelimeler****: veri madenciliği, refah düzeyi, legatum refah endeksi, kümeleme analizi.*

**ANALYSIS OF THE COUNTRIES ACCORDING TO THE PROSPERİTY LEVEL WITH DATA MINING**

***Abstract***

*Data mining (DM) include techniques for meaningful information hidden in these massive data stacks. As an application-oriented field, VM combines many techniques from other fields such as statistics, machine learning, pattern recognition, database and data warehouse systems, information retrieval, visualization, algorithms, high-performance computing, and many application areas. Cluster Analysis (CA), which is one of these techniques, divides the units whose natural groupings are not known precisely, in the raw data matrix into groups (subsets) in terms of their basic characteristics. For this, algorithms gathered in two groups, hierarchical and non-hierarchical, are used. The aim of this study is to divide the countries into groups according to their prosperity levels with CA, which is one of the DM techniques, and to show the applicability of the method. In this context; the 2019 updated data of 12 prosperity indicators and 167 countries in The Legatum Prosperity Index (LPI) were used. In the study; countries were divided into groups with the Wards algorithm, one of the hierarchical CA methods, and the similarities between the countries were determined with the K-Means, one of the non-hierarchical CA methods. Turkey's place in the groups was determined. Results shows that countries are divided into three clusters according to their prosperity levels. The most effective indicators in dividing them into clusters are "market access and infrastructure, education, investment environment", the least effective indicators are "social capital, natural environment, safety and security". On the basis of clusters; it has been determined that the cluster, which includes the countries with the highest welfare level, is stronger than the others in terms of "management, enterprise conditions, investment environment" indicators. The cluster, which includes the countries with the lowest level of welfare, is quite weaker than the others in terms of "education, health, living conditions" indicators. In the middle prosperity level cluster where Turkey is located; it was determined that the "health, living conditions, education" indicators were the highest and the "natural environment, personal freedoms, management" indicators were the lowest. The results reveal that the CA is effective in clustering countries according to the prosperity levels. According to this; in the countries with the highest prosperity level, it can be said that the relationship and harmony between individuals and institutions are strong, and growth is facilitated by innovation, investment, business and trade incentives. In the countries with the lowest prosperity level, it can be said that the aspects that represent social welfare and enable individuals to reach their autonomy and full potential are quite weak. It has been seen that Turkey is at a middle level in matters representing social prosperity, and weak in terms of individuals' basic freedoms and self-development.*

***Keywords:*** *data mining, prosperity level, legatum prosperity index, cluster analysis.*