



## Kümeleme Yaklaşımı ile Maske Üretim Faktörlerinin Analiz Edilmesi Üzerine Bir Uygulama

### An Application on Analysis of Mask Production Factors with Clustering Approach

Semra Tebrizcik<sup>1\*</sup>, Süleyman Ersöz<sup>1</sup>, Adnan Aktepe<sup>1</sup>

<sup>1</sup>Kırıkkale Üniversitesi, Mühendislik ve Mimarlık Fakültesi, Endüstri Mühendisliği Bölümü, Kırıkkale, TÜRKİYE

Başyuru/Received: 22/09/2022

Kabul / Accepted: 30/05/2023

Çevrimiçi Basım / Published Online: 30/06/2023

Son Versiyon/Final Version: 30/06/2023

#### Öz

Teknolojinin gelişmesi ile birlikte büyük veri tabanları daha ulaşılabilir hale gelmiştir. Günümüzde de birçok alanda büyük veri tabanlarının kullanımına imkân verilmektedir. Veri tabanlarında tutulan veriler ile veri madenciliği yaklaşımlarından faydalanarak anlamlı bilgiler ve kurallar keşfedilmektedir. Böylelikle işletmeler için yararlı olacak bilgi keşifleri yapılmaktadır. Bu çalışmada cerrahi (tıbbi) maske üretimi yapmakta olan bir fabrikanın üretim verileri kullanılarak kalite kontrol aşamasında hatalı veya hatasız maskelerin üretilmesinde etkili olan değişkenlerin belirlenmesi doğrultusunda kümeleme analizleri yapılmıştır. Analizler Two-step (İki Aşamalı Kümeleme) yöntemi ile yapılarak ortaya çıkan kümelerin özellikleri karşılaştırılmıştır. Küme bölümlenmesinde en çok etkili olan değişkenlerin; maske gövdesinin oluşturulmasında kullanılan hız miktarı, 3 katmandan oluşan maske gövdesinin orta katmanına ait kumaşın türü ve gövde üzerinde bulunan pilelerin oluşturulması için kullanılan ultrasonik ısı miktarının olduğu gözlemlenmiştir. Üretim faaliyetlerinin performansını artırmak için elde edilen bilgiler süreci iyileştirmek için kullanılacaktır.

#### Anahtar Kelimeler

“Kümeleme analizi, İki Aşamalı Yöntem, Veri Madenciliği, SPSS clementine”

#### Abstract

Big databases have become more accessible with the development of technology. Today, it is possible to use large databases in many areas. By using data kept in databases and data mining approaches, meaningful information and rules are discovered. Thus, information discoveries are made that will be useful for businesses. In this study, using the production data of a factory that produces surgical (medical) masks, clustering analyzes were carried out in order to determine the variables that are effective in the production of defective or defective-free masks at the quality control stage. The analyzes were made with the Two-step method and the properties of the resulting clusters were compared. The most effective variables in cluster partitioning are; It has been observed that the amount of speed used in the creation of the mask body, the type of fabric of the middle layer of the mask body consisting of 3 layers, and the amount of ultrasonic heat used to create the pleats on the body. The information obtained to improve the performance of production activities will be used to improve the process.

#### Key Words

“Cluster analysis, Two-step method, Data mining, SPSS clementine”

## 1. Giriş

Günümüzde bilgi işlem teknolojisinde yaşanan gelişmeler, işletmelerin yönetilmesinde ve etkin karar vermesinde önemli avantajlar sağlamaktadır. Bilgi işlem teknolojilerinin geldiği nokta çok miktarda verileri toplamamıza, kaydetmemize ve saklamamıza imkân vermektedir. Mevcut verilerin analizlerinin yapılarak bilgiye dönüştürülmesi ve süreç ile ilgili kararlar verilmesinde veri madenciliği yaklaşımlarından yararlanılmaktadır.

Üretim faktörleri (öznitelikler, değişkenler) farklı aşamalarda sürekli olarak ölçülmekte ve değerleri işletmelerin veri tabanlarında saklanmaktadır. Bu veriler üretim prosesindeki üretim hattında hangi makinenin hangi kurulum parametreleriyle kullanıldığı, operatörlerin özellikleri (tecrübesi, yaşı, vardiya tipi vb.), proseste kullanılan hammaddeler, ortam (nem, sıcaklık vb.), makinelere takılan sensörler (titreşim, kuvvet, basınç vb.), makine arızaları ve bakımları, ürün kalitesi ve diğer önemli özellikleriyle ilgili olabilmektedir. Bu bağlamda veri madenciliği genellikle veri örneklerinin özellikleri arasındaki karmaşık ilişkileri anlayarak yeni bir örnek için bilinmeyen değişken değerlerini öngörebildiğinden üretimdeki diğer geleneksel matematiksel ve istatistiksel modellerden daha verimlidir (Doğan & Birant, 2021).

Üretim proseslerinde verimliliği etkileyen en önemli faktörlerden biri kalitedir. Üretim işletmelerinde rekabet ortamında ayakta kalabilmek için üretimi gerçekleştirilen ürünün istenilen kalite karakteristiklerine uygun olarak üretilmesi, müşteri memnuniyetinin sağlanması ve üretim maliyetlerinin düşürülmesi gerekmektedir.

Bu çalışmada, COVID-19 salgını ile birlikte faaliyete başlayan cerrahi (tıbbi) maske üretimi yapan bir fabrikanın üretim verileri kullanılarak kalite kontrol aşamasında hatalı veya hatasız maskelerin üretilmesinde etkili olan değişkenlerin kümeleme analizleri yapılmıştır. Kümeleme analizleri sonucu ortaya çıkan kümelerin değerlendirilmesi ile işletme için yararlı olacak bilgi keşfi yapılması hedeflenmiştir. Elde edilen bilgiler üretim faaliyetlerinin performansını artırmak ve süreci iyileştirmek için kullanılacaktır.

Makalenin giriş bölümünü takip eden 2. bölümde literatürde kümeleme yöntemlerini kullanan çalışmalar anlatılmaktadır, 3. Bölümde veri madenciliği kavramı ve kümeleme analizi anlatılmaktadır, 4. bölümde ise çalışmanın uygulama kısmı yer alarak, veri hazırlama süreci, kümeleme analizlerine ve bulgulara yer verilmektedir, 5. Bölümde ise kümeleme analizleri sonuçlarına yer verilerek değerlendirmeler yer almaktadır.

## 2. Literatür Araştırması

Literatürde farklı disiplinlerde kümeleme analizleri ile ilgili pek çok çalışma yapılmıştır.

Helgesen vd. (2009) kümeleme analizi ve regresyonu birleştirerek, uygulayıcıların pazarlama ve işletme performansı algıları arasındaki ilişkileri inceleyerek yöneticilerin pazarlama algılarına göre işletmelerin üç gruba ayrılabilceğini ve bir uygulayıcının ait olduğu kümenin işletme performansında fark yarattığını göstermişlerdir. Çalışmada kümeleme analizleri için En yakın komşu algoritması kullanılmıştır. Savaş & Topaoğlu (2011) GSM şebekelerinin çekim gücü verileri ile iki aşamalı kümeleme (Two-step) tekniği kullanılarak ortaya çıkan 3 kümenin performans analizlerini yapmışlardır. Sonuç olarak A, B ve C şebekelerinin performansları belirlenerek, sinyal gücü temel alındığında A şebekesinin yüksek performans gösterdiği ortaya koyulmuştur. Özdemir & Orçanlı (2011) işletmeler için yararlı olacak pazar bölümlendirmesi üzerinde çalışarak hedef pazarlara dikkat çekmek için kümeleme analizlerinden faydalanmışlardır. Çalışmada Two-step kümeleme yöntemi kullanılarak demografik ve sosyo-kültürel değişkenler ile ortaya 5 adet pazar bölümlenmesi elde edilip her bir kümenin özellikleri açıklanarak yorumlanmıştır. Çetin Irmak (2014) akademisyenlerin elektronik alışverişte güvenlik ve risk algılarını belirlemeye odaklanıp anket çalışmasıyla verileri elde etmiştir. Risk algıları bakımından veriler incelenmek üzere iki aşamalı kümeleme yöntemi kullanılarak kümeleme analizleri yapılmış ve katılımcıların risk algılarına göre 4 farklı kümede toplandığı görülmüştür. Çalışmada kümeler arasında oluşan risk algılarındaki farklılıklar yorumlanmıştır. Ödev ve Peker (2022) internet ortamında firmaların aldığı müşteri şikayetlerini ve bunların yönetim performanslarını değerlendirmek üzere Ward tekniği ve k-ortalama algoritmasını bir arada kullanan, iki aşamalı kümeleme yöntemi analizlerini kullanmışlardır. Analizlerde k küme sayısını belirlemek için ilk olarak Ward'ın hiyerarşik yöntemi kullanılmış ve ardından verileri k gruba ayırmak için k-ortalama kümeleme yöntemi kullanılmıştır. Analizler sonucu ortaya çıkan farklı firma küme özellikleri ile firmaların, kendi performanslarını değerlendirmelerine, kendilerine iyileştirici yeni hedefler belirlemelerine olanak sağlamaktadır.

Zidek vd. (2016) yüksek hacimli imalat ürünlerinde otomatik tanımlama ve sistematik hataların teşhisinde benzer hataların toplanması için kümeleme ve önerilen sınıfa herhangi bir yeni hatanın gelmesi tahmini için sınıflandırma yöntemleri kullanmıştır. Çalışma üç ayrı işleme seviyesinden oluşur: ürünlerdeki hataların belirlenmesi için görüntü elde etme, hataları ayrı sınıflara kategorize etmek için veri kümeleme yöntemlerini kullanma ve önerilen bir sınıf modeliyle yeni tahmin modeli kurmak. Kümeleme analizlerinde En yakın komşu algoritması, k-ortalama ve DBSCAN (Density-Based Spatial Clustering of Applications with Noise) yöntemleri kullanılmıştır. Çalışma & Baynal (2016) bankacılık sektöründe yapmış oldukları çalışmalarında kümeleme yöntemlerinden k-ortalama tekniğini kullanarak bir banka şubesine müşterilerin, belirlenen on iki değişken ile kümelenebilirliği ve bu kümelemede bulunan müşteri profillerine göre satış stratejileri geliştirilmesi amaçlanmıştır. Analiz sonuçlarında çıkan kümelerin özellikleri yorumlanarak, satış stratejisi önerileri verilmiştir. Alsaedi (2017) tez çalışmasında twitter mesajlarını özetlemeye odaklanan teknikler olarak kümeleme analizlerine de yer

vererek On-line Clustering kümeleme yöntemini kullanmıştır. Sınıflandırma yöntemleriyle entegre bir şekilde kurulan çerçevede kümeleme analizleri ile Twitter gönderileri, dikkat edilmesi gereken gönderiler şeklinde ayırt edilmeye çalışılmıştır. Uygulama sonuçları değerlendirildiğinde birçok dilde yazılmış milyonlarca sosyal medya mesajından oluşan büyük veri kümelerini kullanan bazı uygulamalarla karşılaştırılmıştır. Sonuçlar yapılan çalışmanın daha iyi performans gösterdiğini ortaya koymuştur. Analiz sonuçları gerçek dünya olaylarını çevreleyen kamu güvenliğinin güvence altına alınmasında kullanılmıştır. Ceylan vd. (2017) bir emeklilik şirketinin müşterileri hakkındaki bilgilere belirledikleri sekiz değişken ile kümeleme yaklaşımlarından biri olan iki aşamalı kümeleme (Two-step) yöntemi kullanarak ulaşımlardır. Kümeleme analizlerinden elde edilen bulgular şirketlerin yeni satış stratejileri geliştirirken hangi değişkenleri göz önüne alarak yapılması gerektiğini göstermiştir.

Köse (2022) 36 OECD ülkesinin 2019 yılı sağlık araştırma raporunun verilerini kullanarak ülkelerin kümeleme analizleri ile değerlendirmelerini yapmıştır. Analizlerde seçilen sağlık insan gücü, sağlık statüsü, sağlık sektörü kaynakları ile ilgili göstergeler açısından k-ortalamlar kümeleme yöntemine göre nasıl gruplandığı belirlenmiştir. Hernández vd. (2018) alerjik astım rahatsızlığı olan belli bir grup hasta için 19 değişken belirleyerek hastaların benzerliklerine göre analizini Ward kümeleme yöntemini kullanarak gerçekleştirmişlerdir. Analizler sonucu ortaya çıkan üç kümedeki değişkenler ile alerjik astımı olan hastaların özellikleri değerlendirilmiştir. Özkan vd. (2019) nüfus yoğunluğu oldukça fazla olan bir büyükşehir belediyesinin beş büyük ilçesine ait çağrı merkezi verilerini kullanarak, k-ortalamlar yöntemi ile belirlemiş oldukları yedi farklı değişkene göre kümeleme analizleri yapmışlardır. Analizler SPSS Clementine ve WEKA veri madenciliği paket programları kullanılarak yapılmış ve çıkan kümelerin özellikleri karşılaştırılarak yorumlanmıştır. Choi vd. (2021) yaptıkları çalışmanın kümeleme analizlerinin yapıldığı bölümünde bir işletmenin servis hizmeti verirken servis yetkililerinin özellikleri yetenekleri doğrultusunda gruplandırılmışlardır. Kümeleme analizleri için k-ortalamlar yöntemi kullanılmıştır. Sonuçlar parçaların ömrünü uzatarak ve maliyeti düşürerek bir kalite iyileştirmesi elde etmek için önerilebileceğini göstermiştir.

Bu çalışmada da günümüzde önemi artan cerrahi maske sektörüne ait veriler kullanılmıştır. Kalite kontrol aşamasında ortaya çıkan hatalı veya hatasız maskelerin üretilmesinde etkili olan 14 değişken belirlenmiştir. Kümeleme analizlerinde hem sürekli hem de kategorik verilerin kullanılabilmesine imkân verebilen Two-step kümeleme yöntemi kullanılmıştır. Kümeleme analizlerinin yapılması sonucunda kümelerin özellikleri karşılaştırılmıştır.

### 3. Veri Madenciliği

Veri madenciliği veya veri tabanlarında bilgi keşfi verilerden faydalı ve anlaşılır örüntülerin ortaya çıkartılabilmesi için gerçekleştirilen bir süreçtir (Fayyad vd., 1996). 1990'ların ortalarında ortaya çıkan veri madenciliği kavramı günümüzde, Fayyad ve arkadaşları tarafından önerildiği gibi belirli analiz yöntemlerinin kullanılmasından ziyade veri analizi sürecini vurgulayan “Veritabanlarında Bilgi Keşfi” ile eşanlamı hale gelmektedir (Çakır vd., 2019).

#### 3.1 Kümeleme Analizi

Kümeleme, veri madenciliği analizlerinin temel yöntemlerinden biridir (Ching & Ng, 2002). Kümeleme analizlerinde (Cluster Analysis) tahmin edilecek bir bağımlı değişken (sınıf, etiket) değeri bulunmaz, gözlemlerin ya da değişkenlerin sahip oldukları özellikler çerçevesinde kümelemek (homojen alt gruplara ayırmak) amacıyla geliştirilmiş denetimsiz öğrenme yaklaşımı analiz yöntemlerinden biridir. Sınıflandırma algoritmalarından farklı olup benzerliklerine (ilişkilerine) bağlı olarak doğal sınıflandırmalar oluşturulur. Kümeleme analizi sonucunda elde edilecek kümelerin kendi içinde benzer (homojen), kendi aralarında olabildiğince farklı (heterojen) bir yapıda olması beklenir (Packianather vd., 2017). Literatürde yaygın kullanılan kümeleme yaklaşımları, hiyerarşik (aşamalı) ve hiyerarşik olmayan kümeleme yöntemi olmak üzere iki gruba ayrılmıştır (Atbaş, 2008).

#### 3.2 Two-step Yöntemi (İki Aşamalı Kümeleme)

Two-step kümeleme yöntemi, çok büyük boyuttaki veri kümelerini çıkarmak üzere oluşturulmuş ölçeklenebilir bir kümeleme analizi algoritmasıdır (Rundle Thiele vd. 2015). Algoritma ile büyük boyutlu veri setlerinde çalışılabilir, kategorik ve sürekli değişkenlerin bulunduğu veri setleri kullanılabilir. Yöntem optimum küme sayısını belirleyebilir ve elde edilen kümelerle aykırı gözlem değerlerinin istendiğinde veriden ayıklanabilmesine olanak sağlamaktadır (Norusis, 2007).

Two-step kümeleme analizi, ön kümeleme ve kümeleme olmak üzere iki temel aşamadan oluşur (Okazaki, 2007).

- Ön kümeleme aşamasında gözlemler küçük alt kümelere bölünerek ön kümeleme yapılır daha sonra bu alt kümeler ayrı gözlemler olarak ele alınır. Gözlemin daha önceden oluşturulmuş bir kümeleme içerisinde mi yer aldığı yoksa yeni bir kümeleme mi katılacağı kararı uzaklık kriterlerine bağlı olarak hiyerarşik küme yöntemi ile yapılmaktadır.
- Kümeleme aşamasında ise, bir önceki aşamada elde edilen alt kümeler, gerekli küme sayısına göre gruplanır. Two-step kümeleme yöntemi kümelerin sayısını otomatik belirleyen aglomeratif hiyerarşik bir yöntem kullanır. Hiyerarşik kümeleme yöntemi, model kurulumun sonunda tüm gözlemleri içeren bir küme kalana kadar kümelerin ardışık olarak birleştirildiği bir uygulama yöntemidir (Norusis, 2007).

### i. Uzaklık ölçüsü:

Veri seti bir veya daha fazla değişken kategorik bulunduyorsa, olasılığa dayalı mesafeyi temel alan log-olabilirlik uzaklık ölçüsü kullanılmaktadır. Log-olabilirlik uzaklık hesaplamasında, sürekli değişkenler için normal dağılım ve kategorik değişkenler için ise multinomial dağılım varsayımı yapılmaktadır. SPSS Clementine programının, kategorik ve sürekli değişkenlerle uyumlu olmasından dolayı kümeleri birleştirmek için mesafe ölçüsü olarak log-olabilirlik uzaklık ölçüsü kullanılmaktadır (Okazaki, 2007).

- log-olabilirlik uzaklık ölçümü:

$$d(i,j) = \xi_i + \xi_j - \xi(\langle i,j \rangle) \quad (1)$$

burada,  $i$  ve  $j$  kümeleri,  $\xi_i$   $i$  kümesi içinde bir çeşit varyans olmak üzere,  $d(i,j)$   $i$  ve  $j$  kümeleri arasındaki log-olabilirlik uzaklığını,  $\langle i,j \rangle$  ise  $i$  ve  $j$  kümelerini birleştirerek ortaya çıkan kümeyi ifade eder.

$$\xi_i = -N_j \left( \sum_{k=1}^{K^A} \frac{1}{2} \log(\sigma_k^2 + \sigma_{jk}^2) + \sum_{k=1}^{K^B} E_{jk} \right) \quad (2)$$

burada  $j$  küme ve  $k$  değişken (kategorik veya sürekli) olmak üzere  $N_j$   $j$  kümesindeki gözlem sayısını,  $K^A$  toplam sürekli  $K^B$  toplam kategorik değişken sayısını,  $\sigma_k^2$  sürekli değişkeninin varyansını ve  $\sigma_{jk}^2$  sürekli değişkeninin  $j$  kümesindeki varyansını ifade etmektedir.

$$E_{jk} = -\sum_{l=1}^{L^k} \frac{N_{jkl}}{N_j} \log \frac{N_{jkl}}{N_j} \quad (3)$$

Burada ise  $L^k$   $k$ . kategorik değişkenin sınıf sayısını,  $N_j$   $j$ . kümenin gözlem sayısını  $N_{jkl}$  1 kategorili  $k$  değişkeninin bulunduğu  $j$  kümesindeki gözlem sayısını ifade etmektedir.

### ii. Küme Sayısını Belirleme:

Optimum küme sayısı yöntem tarafından otomatik belirlenmektedir. Küme sayısını belirlemede Bayesian Information Criterion-BIC ve Akaike Information Criterion-AIC kullanılmaktadır.

$$BIC(j) = -2 \sum \xi_j + m_j \log(N) \quad (4)$$

$$AIC(j) = -2 \sum_{j=1}^j \xi_j + 2m_j \quad (5)$$

$$m_j = j \left\{ 2K^A + \sum_{k=1}^{K^B} L^k - 1 \right\} \quad (6)$$

## 4. Uygulama

Veri seti içerisinde tespit edilen eksik kayıtlar model çözümlerinin hatalı sonuçlar üretmemesi için bulunduğu özniteliğin istatistiksel özellikleri dikkate alınarak doldurulmuştur. Eksik kayıt bulunan nümerik veri tipine sahip bir öznitelige ait verilerin normal dağılıma sahip olması göz önünde bulundurularak ortalama değeri ile doldurulmuştur. Örneğin veri seti içerisinde, gövde\_boyu (veri tipi nümerik) özniteliği için eksik veriler ortalama değeri olan 16,93 değeri ile doldurulmuştur.

Operatör cinsiyet değişkeninde tek tip değer (kadın) olduğu için tahmin modelinde etkisi olmaması sebebiyle veri setinden çıkarılmıştır. Tutulan veriler içerisinde bazı gözlemlerin sık sık hatalı ürün sınıfına sahip olduğu fark edilerek bu tekrarlar sebep olan durumun arızalı bir makine olduğu tespit edilmiştir. Tutulan kayıtlar içerisinde ilgili makine numarası tespit edilip model eğitiminde yanlışlık oluşturulmaması amacıyla bu makineye ait gözlemler veri setinden çıkarılmıştır.

Kategorik değişkenlerde bulunan eksik verilerde ise en sık tekrar eden (mod) değer eksik kayıtların yerine girilmiştir. Veri seti içerisinde örneğin burun\_teli\_yapisi kategorik değişkenine (1-2) ait eksik kayıtlar mod değeri olan 2 değeri ile doldurulmuştur. Veri dönüştürme işlemlerinde ise bazı nümerik ve string değerler kategorize edilmiştir. Örneğin operatör\_egitim değişkeni; ilkokul=1, lise=2 olacak şekilde 2 sınıfa, oprtr\_yas değişkeni; 18-24, 25-34, 35-44, 45-54 ve 55\_üzeri olacak şekilde 5 sınıfa ve gövde\_üst/orta/alt katmanları için kullanılan kumaş türleri ise A, B, G, M ve T olacak şekilde 5 sınıfa dönüştürülmüştür.

Veri ön işleme ve başlangıçta yapılan veri indirgeme aşamalarından sonra oluşturulan veri seti 14 değişken ve 959 adet gözlemden (satur elemanından) oluşmaktadır. SPSS Clementine ortamında veri setinin ilk 37 gözlem değeri Şekil 4.1'de gösterilmektedir.

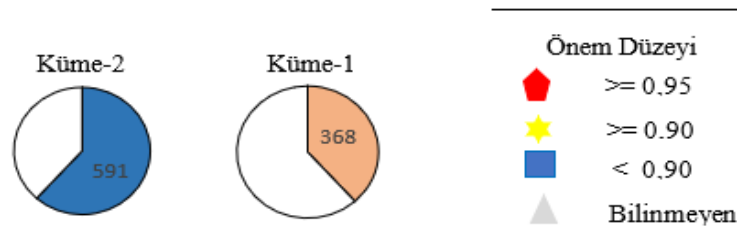
	vardiya	govde_makine_no	govde_makine_hiz_ayari	Oprtr	Oprtr-Yas	Oprtr_1_Egitim	Oprtr_Tecrube	govde_ust_...	govde_orta...	govde_ait_k...	govde_boyu	govde_eni	burun_teli_yapisi	govde_ultrasonik_isisi
1	2.000	1.000	30.000	J	45-54	1.000	9.000	M	B	17.000	9.500	2.000	91.000	
2	2.000	1.000	30.000	J	45-54	1.000	9.000	M	B	17.000	9.500	2.000	91.000	
3	2.000	1.000	30.000	J	45-54	1.000	9.000	M	B	17.000	9.500	2.000	91.000	
4	2.000	1.000	30.000	J	45-54	1.000	9.000	M	B	16.900	9.400	2.000	91.000	
5	2.000	1.000	30.000	J	45-54	1.000	9.000	M	B	17.000	9.500	2.000	91.000	
6	2.000	1.000	30.000	J	45-54	1.000	9.000	M	B	17.000	9.500	2.000	91.000	
7	2.000	1.000	30.000	J	45-54	1.000	9.000	M	B	17.100	9.500	2.000	91.000	
8	2.000	1.000	30.000	J	45-54	1.000	9.000	M	B	17.000	9.500	2.000	91.000	
9	2.000	1.000	30.000	J	45-54	1.000	9.000	M	B	17.000	9.500	2.000	91.000	
10	2.000	1.000	30.000	J	45-54	1.000	9.000	M	B	17.000	9.400	2.000	91.000	
11	2.000	1.000	30.000	J	45-54	1.000	9.000	M	B	17.000	9.400	2.000	91.000	
12	2.000	1.000	30.000	J	45-54	1.000	9.000	M	B	17.100	9.500	2.000	91.000	
13	2.000	1.000	30.000	J	45-54	1.000	9.000	M	B	17.000	9.500	2.000	91.000	
14	2.000	1.000	30.000	J	45-54	1.000	9.000	M	B	17.000	9.500	2.000	91.000	
15	2.000	1.000	30.000	J	45-54	1.000	9.000	M	B	17.000	9.400	2.000	91.000	
16	2.000	1.000	30.000	J	45-54	1.000	9.000	M	B	17.100	9.400	2.000	91.000	
17	2.000	1.000	30.000	J	45-54	1.000	9.000	M	B	17.100	9.500	2.000	91.000	
18	2.000	1.000	30.000	J	45-54	1.000	9.000	M	B	17.000	9.200	2.000	91.000	
19	2.000	1.000	30.000	J	45-54	1.000	9.000	M	B	17.100	9.500	2.000	91.000	
20	2.000	1.000	30.000	J	45-54	1.000	9.000	M	B	17.100	9.500	2.000	91.000	
21	2.000	1.000	30.000	J	45-54	1.000	9.000	M	B	17.100	9.500	2.000	91.000	
22	2.000	1.000	30.000	J	45-54	1.000	9.000	M	B	17.100	9.400	2.000	91.000	
23	2.000	1.000	30.000	J	45-54	1.000	9.000	M	B	17.100	9.500	2.000	91.000	
24	2.000	1.000	30.000	J	45-54	1.000	9.000	M	B	17.000	9.500	2.000	91.000	
25	2.000	1.000	30.000	J	45-54	1.000	9.000	M	B	17.000	9.500	2.000	91.000	
26	2.000	1.000	30.000	J	45-54	1.000	9.000	M	B	17.000	9.600	2.000	91.000	
27	2.000	1.000	30.000	J	45-54	1.000	9.000	M	B	17.100	9.500	2.000	91.000	
28	2.000	1.000	30.000	J	45-54	1.000	9.000	M	B	17.000	9.500	2.000	91.000	
29	2.000	1.000	30.000	J	45-54	1.000	9.000	M	B	17.100	9.400	2.000	91.000	
30	2.000	1.000	30.000	J	45-54	1.000	9.000	M	B	17.000	9.400	2.000	91.000	
31	2.000	1.000	30.000	J	45-54	1.000	9.000	M	B	17.100	9.500	2.000	91.000	
32	2.000	1.000	30.000	J	45-54	1.000	9.000	M	B	17.000	9.400	2.000	91.000	
33	2.000	1.000	30.000	J	45-54	1.000	9.000	M	B	17.000	9.500	2.000	91.000	
34	2.000	3.000	30.000	C	18-24	2.000	9.000	B	G	16.700	9.500	1.000	70.000	
35	2.000	3.000	30.000	C	18-24	2.000	9.000	B	G	16.900	9.500	1.000	70.000	
36	1.000	2.000	50.000	F	45-54	1.000	5.000	B	G	16.800	9.500	2.000	70.000	
37	1.000	2.000	50.000	F	45-54	1.000	5.000	B	G	16.900	9.400	2.000	70.000	

Şekil 4.1 SPSS Clementine Ortamında Veri Setinin Görünümü

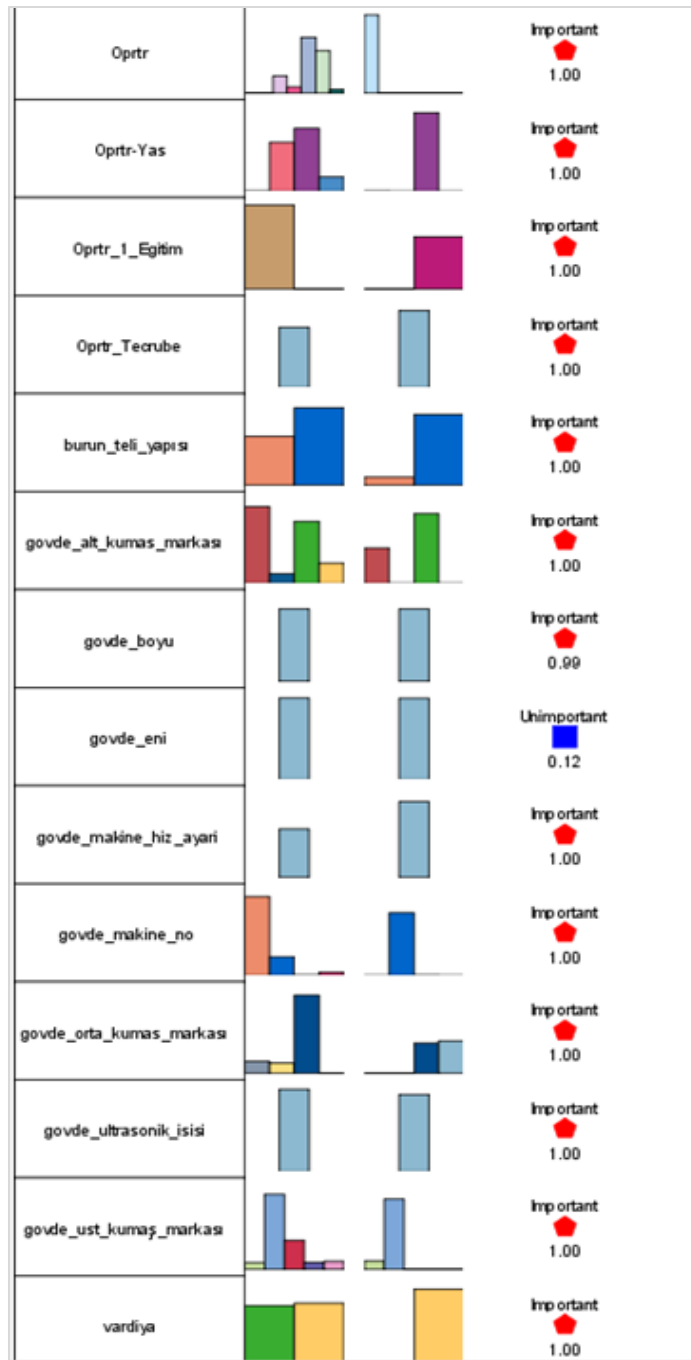
Bu çalışmada, Two-step kümeleme analizi yönteminin kullanılmasının temel amacı ele alınan, hatalı veya hatasız maskelerin üretilmesinde etkili olan değişkenlere ait üretim verilerini, değişkenlere göre aynı kümede toplayarak ortaya çıkan farklı grupları belirlemek ve bu gruplar üzerindeki değişkenlerin önemini gözlemleyebilmektir.

SPSS Clementine 12.0 versiyonu 3 ile 15 küme arasında bir değer aralığı sunarak küme sayısını otomatik vermektedir. Ayrıca specify number of cluster seçeneği ile de kullanıcının istediği küme sayısını belirtmesini sağlamaktadır. Bu çalışmada küme sayısına ilişkin daha önceden bilgi sahibi olduğundan specify number of cluster seçeneği kullanılarak küme sayısı 2 olarak belirtilmiştir. Ortaya çıkan kümelerde küme-1 368, küme-2 591 kayıttan oluşmaktadır.

Şekil 4.2'de kategorik ve nümerik değişkenlerin 2 farklı kümedeki frekans değerleri gösterilmektedir ayrıca veri setindeki her değişkenin önem (important) düzeyleri incelenmiştir. Veri setinde bulunan gövde eni değişkeni %90'dan düşük önem düzeyine sahiptir bunun dışında diğer tüm değişkenlerin önem düzeylerinin % 95 ve üzeri bir değerde olduğu görülmüştür bu da seçilen değişkenlerin kümeleme analizi için önemli değişkenler olduğunu göstermektedir.



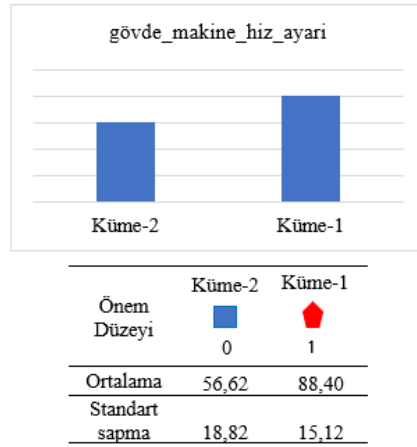
Şekil 4.2 Two-step Kümeleme Yöntemi ile Elde Edilen Kümeler ve Değişkenlerin Önem (important) Düzeyleri



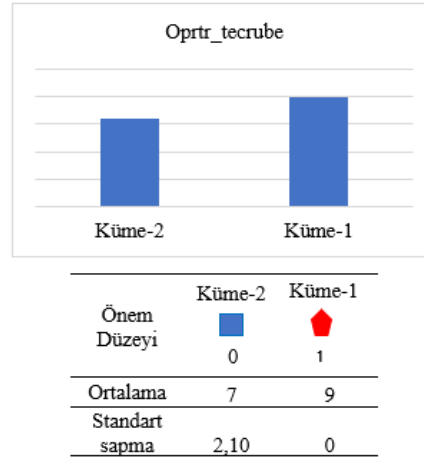
Şekil 4.2 (devam) Two-step Kümeleme Yöntemi ile Elde Edilen Kümeler ve Değişkenlerin Önem (important) Düzeyleri

Şekil 4.3'de makine hız ayarı değişkeni ele alınarak oluşturulan 2 küme değerlendirildiğinde küme-1'de hız değeri ortalaması (mean) 88,40 ve standart sapma değeri (standart deviation) 15,12'dir. Küme-2'de hız değeri ortalaması ise 56,62 ve standart sapma değeri 18,82'dir. Gövde makinesinde işlem gören ürünlerden Küme-1'de bulunan ürünlerin maruz kaldığı hız ortalama 88,40 ve Küme-2'deki ürünlerin ortalama 56,62 değerinde bir hız miktarı ile işlem gördüğünü göstermektedir. Bu değişken kümelerin farklılaşmasında etkili bir değişkendir.

Operatör tecrübesi değişkeni ele alınarak oluşturulan (Şekil 4.4) 2 küme değerlendirildiğinde Küme-1'de tecrübe ortalaması 9 (std:0)'dur. Küme-2'de tecrübe ortalaması 7 (std:2,10)'dir. Operatör tecrübesi değişkeninin yüksek olması dolayısıyla Küme-1 çoğunlukla hatasız ürünlerin gruplandığı küme olarak değerlendirilebilir. Bu değişken kümelerin farklılaşmasında etkili bir değişkendir.

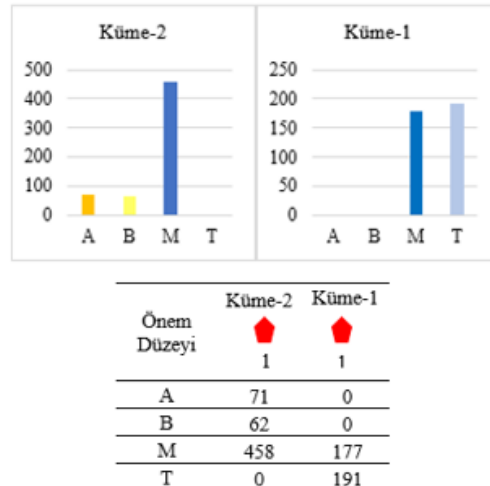


Şekil 4.3. Gövde Makine Hızı Değişkenine Göre Küme-1 ve Küme-2



Şekil 4.4 Operatör Tecrübesi Değişkenine Göre Küme-1 ve Küme-2

Şekil 4.5'deki sonuçlar incelendiğinde ise kullanılan kumaş türlerinden M türü Küme-2'de daha fazla sayıya (458 adet) sahiptir. Küme-1'de ise M (177 adet) ve T (191 adet) türüne ait kumaş türlerinin daha fazla sayıda olduğu görülmüştür. Ayrıca T türü kumaşın Küme-2'de yer almaması ve Küme-1'de ise diğer kumaş türlerine göre en fazla sayıda yer alması T kumaş türünün 2 kümenin farklılaşmasında etkili olduğunu gösterir. Çıktı olarak elde edilen sonuçlar Tablo 4.1 ve Tablo 4.2'de veri tiplerine göre gösterilmiştir.



Şekil 4.5 Gövde Orta Kumaş Türü (markası) Değişkenine Göre Küme-1 ve Küme-2

**Tablo 4.1** Nümerik Değişkenlerin Küme-1 ve Küme-2'deki Ortalama Değerleri

	Govde_makine_hızı	Oprt_tecrube	Govde_ultrasonik_isisi	Govde_boyu	Govde_eni
<b>Küme-1</b>	88,40	9	72,20	16,92	9,49
<b>Küme-2</b>	56,62	7	77,08	16,94	9,52

**Tablo 4.2** Kategorik değişkenlerin Küme-1 ve Küme-2'deki Frekans Değerleri

Üretim Faktörleri (Kategorik Değişkenler)									
	Govde_orta_kumaş_turu	Govde_ust_kumaş_turu	Govde_alt_kumaş_turu	Oprt_egitim	Oprtr_yas	Burun_teli_yapisi	vardiya	Govde_makine_no	oprt
<b>Küme-1</b>	M (177)	A (39)	A (269)		18-24				
	T (191)	B (329)	B (33)	2 (368)	(1)	1 (38)	2 (368)	2 (367)	B (367)
<b>Küme-2</b>			G (218)		45-54	2 (330)			C (1)
			T (71)		(367)				
		A (31)			35-44				
	A (71)	B (352)	A (124)	1 (590)	(229)	1 (229)	1 (290)	1 (462)	F (80)
	B (62)	G (137)	G (244)	2 (1)	45-54	2 (362)	2 (301)	2 (109)	G (29)
M (458)	M (33)			55_uzeri			3 (1)	J (262)	
	T (38)			i (67)			6 (19)	L (200)	
								M (19)	

Nümerik değişkenlerin Küme-1 ve Küme-2'deki ortalama değerlerine dikkat edildiğinde gövde\_makine\_hızı Küme-1'de 88,40 Küme-2'de 56,62'dir. Bu durum 2 kümenin ayrışmasında kullanılan değişkenin önemli olduğunu göstermektedir. Gövde boyu ve gövde\_eni değişkenlerinin her iki kümede yaklaşık aynı ortalamaya sahip oldukları ve kümeleri ayırıcı bir değişken olmadığı söylenebilir. Kategorik değişkenlerin Küme-1 ve Küme-2'deki sınıflarına ait frekans değerlerine dikkat edildiğinde operator eğitim değişkeninin 1 sınıfının Küme-2'de, 2 sınıfının ise Küme-2'de gruplandığı gözlemlenmektedir. Bu farklılık 2 kümenin ayrışmasında kullanılan değişkenin önemli olduğunu göstermektedir. Gövde orta kumaş türü değişkenine bakıldığında ise M sınıfına ait kumaşların ağırlıklı olarak Küme-2'de gruplandığı gözlemlenmektedir.

## 5. Sonuç

Çalışmada COVID-19 salgını ile birlikte tüm dünyada kullanımı yaygınlaşan koruyucu ekipman olarak kullandığımız cerrahi (tıbbi) maskenin gövde üretim prosesindeki üretim faktörleri kullanılarak kümeleme analizi yapılmıştır. Veri ön işleme aşamasından sonra elde edilen değişkenlere ait üretim verileri kullanılarak hatalı veya hatasız 2 sınıftan oluşan veri setine veri madenciliği yaklaşımlarından kümeleme analizi (denetimsiz öğrenme) uygulanmıştır.

Kümeleme uygulaması için kategorik ve sürekli değişkenlerin bulunduğu veri setlerine uygun olan Two-step algoritması kullanılmıştır. Elde edilen veriler ışığında küme bölümlenmesinde en çok etkili olan değişkenlerin maske gövdesinin oluşturulmasında kullanılan hız miktarı, 3 katmandan oluşan maske gövdesinin orta katmanına ait kumaş türü ve gövde üzerinde bulunan pilelerin oluşturulması için kullanılan ultrasonik ısı miktarı olduğu belirlenmiştir. Elde edilen bilgiler üretim faaliyetlerinin performansını artırmak ve süreci iyileştirmek için kullanılacaktır. Bu çalışmanın sınırları, uygulamanın sadece gövde üretim prosesinde yapılmış olmasıdır. Uygulamanın yapıldığı işletmede prosesler farklı atölyelerde yürütüldüğünden bir ürüne ait değişkenlerin tamamının takibi mümkün olmadığı için proses baştan sona uygulamaya alınmamıştır.

Gelecek çalışmalarda hatalı maskelerin ortaya çıkmasında etkili olan değişkenlere ait veriler kullanılarak bunlarla birlikte hata türleri belirlenip analizler yapılabilir. Bu yolla hata türleri için gerekli olan proaktif yaklaşımlar ve iyileştirmeler uygulanabilir.

## Referanslar

Alsaedi, N., (2017). Event identification in social media using classification-clustering framework, Doktora Tezi, Cardiff University, 2017.



Aslı, K. Ö. S. E. (2022). OECD ülkelerinin sağlık göstergeleri açısından kümeleme analizi yöntemine göre değerlendirilmesi. Nevşehir Hacı Bektaş Veli Üniversitesi SBE Dergisi, 12(4), 2010-2021.

Atbaş, A., (2008). Kümeleme analizinde küme sayısının belirlenmesi üzerine bir çalışma, Yüksek Lisans Tezi. Ankara Üniversitesi, Ankara.

Ceylan, Z., Gürsev, S. & Bulkan S., (2017). İki Aşamalı Kümeleme Analizi ile Bireysel Emeklilik Sektöründe Müşteri Profiline Değerlendirilmesi, Bilişim Teknolojileri Dergisi, 10(4), 475-485.

Ching W.K., & Ng, K. M., (2002). Advances in Data Mining and Modeling, World Scientific, 1st ed., Hong Kong, China.

Choi, Y., Lee, H., Yang, J., (2021). Development of a service parts recommendation system using clustering and classification of machine learning, Expert Systems With Applications, 188, 1-12.

Çakır, A., Çalis H. & Küçüksille, Ecir U. (2019). Data mining approach for supply unbalance detection in induction motor, Expert Systems with Application, 36, 11808–11813.

Çalış A. & Baynal K., (2016). Kümeleme Analizi İle Bankacılık Sektöründe Satış Stratejilerinin Belirlenmesi, Beykent Üniversitesi Fen ve Mühendislik Bilimleri Dergisi, 9(1), 13 – 41.

Çetin H., & Irmak, S., (2014). Elektronik Alışverişte Akademisyenlerin Güvenlik ve Risk Algılarının Belirlenmesi, Yönetim Bilimleri Dergisi, 12(24), 275-294.

Dogan, A. & Birant D., (2021). Machine learning and data mining in manufacturing, Expert Systems With Applications, 166, 114060.

Fayyad, U., Piatetsky Shapiro G. & Smyth, P., (1996). From data mining to knowledge discovery in database, AI Magazine, 17, 37-54.

Gamze, Ö. D. E. V., & Peker, S. Müşteri şikâyet yönetiminde firmaların performanslarının değerlendirilmesi: Kümeleme analizi incelemesi. Dicle Üniversitesi Mühendislik Fakültesi Mühendislik Dergisi, 13(3), 447-456.

Helgesen, O., Nettet, E. & T. Voldsund, (2009). Marketing perceptions and business performance: Implications for marketing education?, Marketing Intelligence Planning, 27, 25-47.

Norusis, M. J. (2007). SPSS 15.0 advanced statistical procedures companion, Chicago, IL: Prentice Hall.

Okazaki, S., (2007). Lessons learned from i-mode: What makes consumers click wireless banner ads?, Computers in Human Behavior, 23, 1692–1719.

Özdemir, A. & Orçanlı, K., (2012). İki Aşamalı Kümeleme Algoritması ile Pazar Bölümlemesi, Müşteri Profillerinin Belirlenmesi ve Niş Pazarların Tespiti, Uşak Üniversitesi Sosyal Bilimler Dergisi, 5(3), 1-27.

Özkan, E., Avcı, S. & Aladağ, Z., (2019). Büyükşehir Belediyesi Çağrı Merkezi Verilerinin Kümeleme Analizi ile İncelenmesi, Erciyes Üniversitesi Fen Bilimleri Enstitüsü Dergisi, 35(3), 78-91.

Packianather, S., Michael Alan, D., Harraden, S., Soman, S. & White, J., (2017). Data Mining Techniques Applied to a Manufacturing, SME. Procedia CIRP, 62, 123–128.

Rundle-Thiele, S., Kubacki, K., Tkaczynski A. & Parkinson, J. (2015). Using two-step cluster analysis to identify homogeneous physical activity groups, Marketing Intelligence and Planning, 33, 522-537.

Savaş, S. Nettet, E. & Topaoğlu N., (2011). Veri Madenciliği Yöntemi İle GSM Şebekelerinin Performans Analizi, Gazi Üniversitesi Mühendislik Fakültesi Dergisi, 26(4), 741-751.

Sendín-Hernández, M. P., Ávila-Zarza, C., Sanz, C., MarcosVadillo, A., Muñoz-Bellido, E. & Dávila, F. J., (2018). Cluster analysis identifies 3 phenotypes within allergic asthma, The Journal of Allergy and Clinical Immunology: In Practice, (6) 955–961.

Zidek, K., Maxim, V., Pitel, J. & Hosovsky, A., (2016). Embedded vision equipment of industrial robot for inline detection of product errors by clustering-classification algorithms, *Int. J. Adv. Robot. Syst.*, 13, 1-10.