

## COVID-19 Pandemisinde Semptomlar ve Vaka Profiline Veri Madenciliği Yaklaşımıyla İncelenmesi

Gökçe KARAHAN ADALI

Dr. Öğr. Üyesi, Haliç Üniversitesi Yönetim Bilişim Sistemleri Bölümü, İstanbul, Türkiye

**Orcid Numarası:** 0000-0001-8567-4626

**Geliş Tarihi:** 29.07.2022

**\*Sorumlu Yazar e mail:** gokceadali@halic.edu.tr

**Kabul Tarihi:** 07.09.2022

**Atf/Citation:** Karahan Adalı, G. (2022). "COVID-19 Pandemisinde Semptomlar ve Vaka Profiline Veri Madenciliği Yaklaşımıyla İncelenmesi", *Haliç Üniversitesi Sosyal Bilimler Dergisi* 2022, 5(2). 253-269

### Öz

Tüm dünyaya hızla yayılan Korona virüsü 2019 (COVID-19), SARS ve MERS salgınları sonrası görülen üçüncü koronavirüs salgını olarak tüm dünyayı etkisi altına almıştır. Virüs dünya çapında halk sağlığı için büyük bir tehdit olmaya devam etmektedir. Dünya Sağlık Örgütü tarafından pandemi ilan edilen bu salgın, insanların sağlığına, eğitimine, sosyalleşmelerine ve küresel ekonomiye çok büyük etkiler bırakan önemli bir halk sağlığı krizidir. COVID-19'a daha hızlı yanıt verebilmek için her yönüyle çalışmalar devam etmektedir. Bu kapsamda hastalığın seyrinde, hastaların benzer özellikleri taşıdıkları semptomlar ve iyileşme sürecine etkileri önem arz etmektedir. Bu çalışmanın amacı, COVID-19 hastaları arasında yaş, cinsiyet, hastalığın seyrinde görülen semptomların, iyileşme ve ölüm durumu üzerinde nasıl etkili birliktelikler meydana getirdiğini ortaya çıkartmaktır. Çalışmada, sık görülen semptomları belirlemek ve demografik verilerin iyileşme sürecine etkilerini çıkarmak için birliktelik kuralı adı verilen ve yaygın olarak kullanılan veri madenciliği tekniği R programlama dili kullanılarak uygulanmıştır. Elde edilen kurallarda, salgına yakalanmada yaş faktörünün büyük ölçüde etken olduğu tespit edilmiştir. En belirgin semptom olarak ateş ortaya konmuştur. Ateş semptomunu takiben öksürük semptomunun hastalar arasında sıklıkla rastlandığı görülmektedir. Analizler sonucunda, ölüm oranlarında yaşlı bireylerin gençlere, erkeklerin ise kadınlara göre daha çok kayıp verdiği tespit edilmiştir.

**Anahtar Kelimeler:** Veri Madenciliği, Birliktelik Analizi, Apriori, COVID-19.

# Analysis of Symptoms and Case Profile in the COVID-19 Pandemic Through Data Mining Approach

## Abstract

The corona virus 2019 (COVID-19), spreading all over the globe, has affected the whole world as the third coronavirus epidemic after the SARS and MERS. The virus remains a major threat to public health worldwide. This epidemic, which has been declared a pandemic by the World Health Organization, is a significant public health crisis that has profound effects on people's health, education, socialization, and the global economy. In order to respond more quickly to COVID-19, studies still continue to be conducted in all aspects. In this context, the symptoms whose characteristics go similar among patients, and their effects on the healing process are important in the course of the disease. This study aims to reveal the effective associations among COVID-19 widely used called 'the association rule' was applied using the R programming language in order to identify the common symptoms and to determine the effects of demographic data on the recovery process. In the rules obtained, it has been determined that age factor holds a major role in being affected by the epidemic. Fever has been found to be the most prominent symptom which is followed by the frequently encountered cough symptom among the patients. As a result of the analysis, it has been found that elderly individuals score higher than youth and men score higher than women regarding mortality rates.

**Keywords:** Machine Learning, Data Mining, Association Analysis, Apriori, COVID-19.

## 1. Giriş

Çin'in Hubei eyaletinin başkenti Wuhan'da Aralık 2019'da başlangıçta bilinmeyen bir nedenden dolayı yerel bir pnömöni salgını tespit edilmiş ve hızla bu salgının, ciddi akut solunum sendromu koronavirüsün (SARS-CoV-2) neden olduğu tespit edilmiştir (Dong, Du ve Gardner, 2020: 533-534). Salgın o zamandan beri 208 ülkeye yayılmıştır.

Dünya Sağlık Örgütü (World Health Organization -WHO) tarafından yapılan açıklamada, Temmuz 2022 itibari ile dünya genelinde gözlemlenen bu salgında 565.207.160 'dan fazla vaka tespit edilmiştir

(Ji, Ma, Pepelenbosch ve Pan, 2020: 480). COVID-19 nedeniyle ölüm ile sonuçlanan vaka sayısı 6.373.739'a ulaşmıştır.

Türkiye'de ortaya çıkan vaka sayısı ise resmi rakamlara göre 15.524.071'dir. Ölüm ile sonuçlanan vaka sayısı ise 99.184'e ulaşmıştır (Worldometers, 2022). Dünyayı etkisi altına alan ve milyonlarca can kaybının yaşandığı COVID-19, WHO tarafından 11 Mart 2020'de pandemi olarak ilan edilmiştir. Pandemi; bir kıta, hatta tüm dünya yüzeyi gibi çok geniş bir alanda yayılan ve etkisini gösteren salgın hastalıklara verilen addır. WHO tanımlamasına göre bir pandemi ancak aşağıdaki üç koşulu sağladığında başlamış sayılmaktadır (WHO, 2020):

- Nüfusun daha önce maruz kalmadığı bir hastalığın ortaya çıkışı,
- Hastalığa sebep olan etmenin insanlara bulaşması ve tehlikeli bir hastalığa yol açması,
- Hastalık etmeninin insanlar arasında kolayca ve devamlı olarak yayılması,

Bir hastalık veya tıbbi durum sadece yaygın olması ve çok sayıda insanın ölümüne yol açması nedeniyle pandemi olarak nitelendirilemez, aynı zamanda bulaşıcı olması gerekmektedir. Örneğin, kanser, insanlarda çok sayıda ölüme sebep olan bir hastalık olmasına rağmen bulaşıcı olmadığı için pandemi olarak adlandırılmamaktadır (WHO, 2020). COVID-19 salgını öncesinde, 2000'li ve 2010'lu yıllarda yaşanmış SARS (Şiddetli Akut Solunum Sendromu) ve MERS (Orta Doğu Solunum Sendromu) salgınları da diğer koronavirüsler gibi zoonotik kökenlidir; SARS'ın egzotik hayvanlardan, MERS'in ise develerden insana geçerek epidemilere yol açtığı tespit edilmiştir (Er, Ünal, 2020: 2-4).

Mevcut COVID-19 salgını, SARS (2002-2003) ve MERS (2012-devam eden) salgınları ile hem benzerlik hem de farklılıklar göstermektedir.

3 viral enfeksiyon da genellikle yaşlı, kronik rahatsızlığa sahip olan kişilerde, solunum yolu hastalığına yol açan ateş ve öksürük belirtileri ile ortaya çıkmaktadır. WHO, SARS salgını Temmuz 2003' de ilan etmiştir. Toplam 8096 SARS vakasından 29 ülkede 774'ü ölümlerle sonuçlanmıştır. %9,6'lık bir ölüm oranına ulaşmıştır. MERS hala kontrol altında değildir ve bugüne kadar 27 ülkede %34,4'lük bir ölüm oranı ile 2494 onaylı vaka ve 858 ölümlerle sonuçlandığı raporlanmıştır (Wu ve McGoogan, 2020: 3-6).

Dünyada yaşanan salgınlarda ölüm oranlarının bu kadar yüksek olması, salgına yol açan hastalıkların anlaşılmasını ve koruyucu tedbirlerin alınarak tedaviye yönelik çalışmaların hızlandırılmasını zorunlu kılmaktadır. Bu kapsamda; bu çalışmada 38 ülkeyi kapsayan ve COVID-19 virüsüne yakalanmış hastalara ait bilgilerin yer aldığı veri seti üzerinde, veri madenciliği danışmansız öğrenme yöntemlerinden birliktelik kuralları uygulanarak, hastalığın seyrettiği kişilerde var olan anlamlı birlikteliklerin ortaya konması amaçlanmıştır.

## 2. Literatür Çalışması

COVID-19 ile ilgili yapılan çalışmalar literatürde hızla yerini almaktadır. Bu bölümde benzer yöntemler kullanılarak bu alanda yapılan çalışmalara yer verilmiştir. Bu çalışmada kullanılan veri setinin de içinde yer aldığı araştırmada, enfeksiyonun yayılmasını en iyi şekilde izlemek ve tahmin etmek için ortaya çıkan salgınlar sırasında epidemiyolojik verilere ihtiyaç olduğu vurgulanmaktadır (Bo Xu vd., 2020:1-6). COVID-19 mortalitesi ile sağlık hizmeti kaynaklarının mevcudiyeti arasındaki potansiyel ilişkiyi ölçümlemek adına yapılan çalışmada ölüm oranlarının sağlık bakım yükü ile orantılı olduğu sonucu elde edilmiştir (Ji vd., 2020).

Salgın ile ilgili turizm alanında gerçekleştirilen bir çalışmada, turistlerin ziyaret etmeyi düşündükleri bölgedeki virüsün varyantları ile

oldukları aşının türünü dikkate alarak, destinasyon seçimlerine ilişkin kararlarında rehberlik etmesi amacı ile bu iki değişken arasındaki birliktelik ortaya konmuştur (Atsa'am ve Wairo, 2022: 1537-1538).

Birliktelik kurallarını kullanarak gerçekleştirilen çalışmalarda, COVID-19 virüsünün seyrinde sık görülen bulgular olarak öksürük, septik şok ve solunum sıkıntısı sendromunu işaret etmişlerdir (Tandan vd., 2021:5-7). Wojtusiak çalışmasında büyük bir devlet üniversitesi popülasyonu için günlük bildirilen COVID-19 semptomlar ile bireysel hareketler arasındaki ilişkiyi incelemiştir (Wojtusiak vd., 2021: 114-131). Verilerin analizi sonucunda, baş ağrısının en sık bildirilen semptom olduğunu ve bazı bireylerin pandemi sırasında hareketlerini kısıtlamalarına rağmen, genel popülasyonunun hareketlerini önerildiği gibi değiştirmedeğini görülmüştür.

Kişilerin COVID-19 öncesi ve sonrası alışveriş eğilimleri ortaya konulmaya çalışılmıştır. Gerçekleştirilen pazar sepeti analizi sonucunda, COVID-19 öncesi ve sonrası tüketilen gıda ürünlerinde farklılıklar görüldüğü ve pandemi döneminde müşterilerin eğilimlerinin değiştiği gözlemlenmiştir. Diğer bir çalışmada, Çin'de ilerlemiş olan bitkisel ilaç endüstrisine COVID-19 salgını boyunca da sıklıkla başvurulduğu görülmüştür (Alanlar, 2021-114-116). Geleneksel Çin Tıbbında sayısız bitki arasından en etkili bitkisel çifti bulmak için birliktelik yöntemlerine başvurmuş ve 32 farklı kural ile farklı bitki kombinasyonları elde etmiştir (Xia vd., 2021: 16-25). Çalışmalarda apriori algoritması ile hesaplanan birliktelik kurallarının COVID-19 tanısında kullanılabileceği göstermiştir (Çelik, 2020:121-131). İlgili çalışmada, hastalık semptomlarının birliktelik oranları belirlenmiştir. Belirlenmiş olan bazı hastalık kodlarının temsil ettiği hastalık semptomları bir arada olduğunda bu semptomları taşıyan hastaların %100 kesin olarak COVID-19 hastalığına yakalandığı gözlemlenmiştir.

### 3. Malzeme ve Yöntem

Çalışmanın bu bölümünde, veri madenciliğinde birliktelik kurallı yöntemleri COVID-19 testi pozitif çıkan hastalara ait veri seti üzerinden geliştirilen uygulamaya yer verilmiştir. Uygulamada Veri Madenciliği için Çapraz Endüstri Standard Süreç Modeli (Cross-Industry Standard Process for Data Mining-CRISP) (Shearer, 2000) metodolojisi esas alınmıştır. CRISP Modeli, Problemin Tanımlanması, Veriyi Anlama, Veri Hazırlama, Modelleme, Modelin Değerlendirme ve Seçimi ve Modelin Uygulamaya Geçirilmesi olmak üzere altı aşamadan oluşmaktadır.

#### 3.1. Problemin Tanımlanması

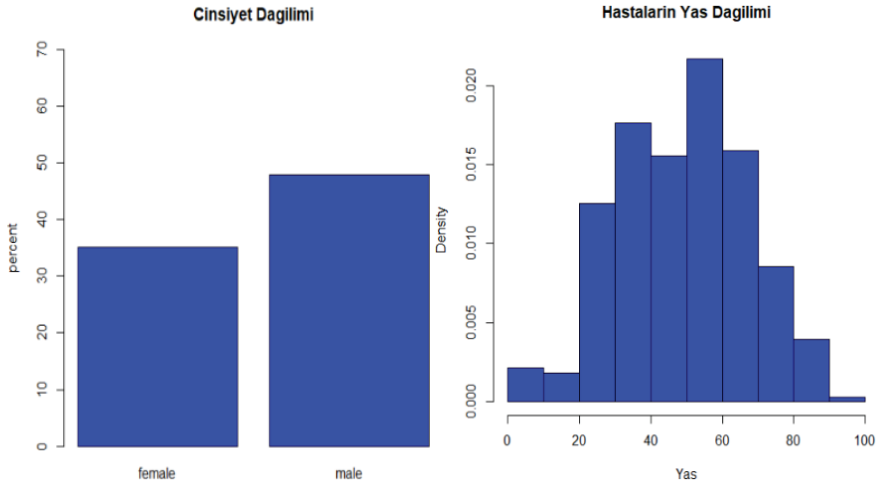
Çalışmada COVID-19 salgınında, hangi semptomların sıklıkla görüldüğü tespit edilmeye çalışılmıştır. Bu kapsamda, veri madenciliği yöntemlerinden birliktelik analizi tekniklerinden apriori algoritması kullanılarak ilgili faktörler arasında en etkili ikili, üçlü birliktelikler ortaya konulmaya çalışılmıştır. Problemin amacı hastalığa yakalanmış kişiler arasında en belirgin özelliklerin neler olduğu tespit etmektir. Analizler sonucunda, hastalığın ölüm ya da iyileşme ile sonlanmasında etken en temel değişkenlerin (cinsiyet, semptom, yaş, vb.) neler olduğu ortaya konmuştur.

#### 3.2. Veriyi Anlama

Çalışmada kullanılacak veri setine kaggle.com sitesi üzerinden erişilmiştir. Veri setinde dünya genelinde COVID-19'dan etkilenen vakalar ile ilgili detaylı bilgilere yer verilmektedir. Veri seti 21 değişken ve 2030 örnekten oluşmaktadır (Kaggle, 2019). Çalışma virüsün çıkış merkezi olan Çin'in Wuhan Eyaleti merkezli gerçekleştirildiğinden, değişkenlerde Wuhan'da bulunma, Wuhan'ın yerel halkı olup olmadığını belirten kategorik değişkenler bulunmaktadır. Hastalığın

seyrinde gözlemlenen semptomlar tek bir sütunda birlikte verilmiş olup, ön işleme ile bu semptomlar birliktelik analizinde kullanılabilir şekilde yan yana ve ayrı hücrelerde bulunacak hale getirilmiştir. Analize etkisi olmayacağı düşünülen açıklama, özet ve kaynak bilgilerini içeren sütunlar analize dahil edilmemiştir.

Çalışma dahilinde eksik veriler çıkarılmıştır. Bunun sonucunda 18 değişken ve 1085 örnek ile analiz gerçekleştirilmiştir. Veri setinde yer alan hastaların cinsiyet dağılımlarına Şekil 1’de, yaş dağılımlarına ise Şekil 2’de yer verilmiştir.



**Şekil 1.** Hastaların Cinsiyet Dağılımı.

**Şekil 2.** Hastaların Yaş Dağılımı.

Hastaların cinsiyet dağılımları histogram üzerinde incelendiğinde, veri setinde yer alan erkek hastaların kadınlara göre daha fazla olduğu görülmektedir. Vakalardaki yaş dağılımlarına bakıldığında, 50-60 yaş grubunun hastalığa yakalanmasının birinci sırada olduğu, bu grubu 30-40 yaş grubunun takip ettiği görülmektedir.

### 3.3. Verinin Hazırlanması

Analizde kullanılan veri setinde yer alan değişkenlere aşağıdaki tabloda yer verilmiştir. Veri seti eksik değerlerin çıkarılması sonrası, toplam 1085 örnekten oluşmaktadır.

**Tablo 2.** Veri setinde yer alan değişkenler.

Değişken Adı	Açıklama	Değişken Adı	Açıklama
1. id	Bildirilen vaka için tanımlayıcı özellik	10.symptom_on_set	Semptom başlangıç tarihi
2.case_in_country	Ülkedeki vaka sayısı	11.if_onset_approxiated	Yaklaşdırılmış başlangıç tarihi (kesin tarih bilgisi ile karşılaştırıldığında)
3. reporting date	Vakanın bildirilme tarihi	12. hosp_visit_date	Vakanın hastaneye geliş tarihi
4. summary	Vaka hakkında özet bilgi	13. exposure_start	Maruz kalma süresinin başlangıcı
5. location	Lokasyon	14. exposure_end	Maruz kalma süresinin bitimi
6. country	Ülke	15. visiting Wuhan	Hastanın Wuhan'ı ziyaret edip etmediği (0=evet, 1=hayır)
7. gender	Cinsiyet	16. from Wuhan	Hastanın Wuhan şehri ile kaydedilmiş ilişkisi (0=Wuhan'da yaşıyor, 1=Wuhan'da yaşamıyor)
8. age	Yaş	17. death	Vefat
9. symptom	Semptom	18. recovered	Kurtarıldı

### 3.4. Modelleme

Çalışmada, birliktelik kuralları adı altında yer alan Apriori algoritmasından faydalanılmıştır. Analiz R programlama dilinde RStudio (RStudio, 2022) geliştirme ortamı yardımıyla gerçekleştirilmiştir. Analiz sırasında birliktelik kurallarına erişebilmek için arules (Haschler



vd., 2022), kuralların görselleştirilmesi için arulesViz (Haschler, 2021) kütüphanesinden faydalanılmıştır. R açık kaynak kodlu bir yazılım olup, istatistiksel hesaplama ve grafikleri için kullanılan programlama dilidir (rproject-2022).

## **Birliktelik Kuralları**

Birliktelik kuralları analizi, yaygın olarak kullanılan veri madenciliğinin tanımlayıcı yöntemlerinden birisidir. Veriler arasındaki birlikteliklerin, ilişkilerin ve bağıntıların kurallar halinde bulunması işlemidir. Birliktelik kuralları, işlemlerden (transaction) oluşan ve her bir işlemin de elemanlarının birlikteliğinden oluştuğu düşünülen bir veri tabanında, bütün öğeleri tarayarak, sık tekrarlanan birliktelikleri veri tabanında ortaya çıkarmaktır (Agrawal ve Srikant, 1996: 487-499).

Veri tabanı kayıtlarında, nesnelerin gruplandırılması ile elde edilen bağımlılık ilişkilerinin yüzde yüz geçerli olması beklenemez. Ancak, çıkarsaması yapılan kuralın, veri tabanının önemli bir kısmı tarafından desteklenmesi, yani söz konusu durumun sıkça görülüyor olması gerekmektedir. Burada,  $A \subset I$ ,  $B \subset I$  ve  $A \cap B = \emptyset$  şeklindedir.  $A \Rightarrow B$  kuralı için, tüm işlem hareketlerinin tutulduğu  $D$  kümesinde yer alan öğeler için  $s$ ; destek değeri olarak tanımlanmaktadır (Agrawal ve Srikant, 1994: 487-499).

Destek değeri (support),  $A$  ve  $B$  nesnelerinin birlikte bulunduğu örnek sayısının, veri tabanındaki tüm örnek sayısına oranı olarak hesaplanmaktadır (Birant vd., 2010: 215-222). Bu kriter, veride öğeler arasındaki bağıntının ne kadar sık olduğunu belirtmektedir.

İkinci olarak,  $A \Rightarrow B$  kuralının  $c$  ile gösterilen güven (confidence) değeri tanımlanmaktadır. Bu kriter çıkarımın (kuralın) gücünü ve güvenilirliğini temsil etmektedir (Nahar vd., 2013: 335-351).  $P(B | A)$  şartlı olasılığı olarak ele alınmaktadır.  $A$  ve  $B$  nesnelerinin birlikte

bulunduğu kayıt sayısının, A nesnesinin (veya nesnelерinin) bulunduğ u kayıt sayısına oranıdır. Destek ve güven değ erlerinin hesaplanmasına Tablo 1’ de yer verilmiştir.

**Tablo 1.** Birliktelik Kurallarında Destek ve Güven Ölçütlerinin Hesaplanması.

$$\begin{aligned} \text{Destek } (A \Rightarrow B) &= P(A \cup B) \\ \text{Güven } (A \Rightarrow B) &= P(B | A) \text{ ya da} \\ \text{Güven } (A \Rightarrow B) &= \text{Destek}(A \Rightarrow B) / \text{Destek}(A) \end{aligned}$$

Kuralın destek ve güven değ erleri, kuralın ilginçliğini ifade eden iki ölçüdü r. Bu değ erler sırasıyla keş fedilen kuralların yararlılığını (kullanış lılığını) ve kesinliğini (doğ ruluğunu) ifade etmektedir (Han ve Kamber, 2006: 243-254). Destek ve güven değ erlerinin eş ik değ eri kullanıcı tarafından belirlenmektedir. Ç ıkarılan kuralın geçerliliğ i aş ağıdaki şartları sağ lıyor ise onaylanmaktadır (Zhang ve Zhang, 2007:1-26).

$$\text{Destek } (X \Rightarrow Y) > \text{mindestek} \quad (1)$$

$$\text{Güven } (X \Rightarrow Y) = \frac{\text{destek } (x \Rightarrow y)}{\text{destek}(x)} \geq \text{min } (\text{güven}) \quad (2)$$

Genel olarak bu ölçütler, 0-1 aralığ ından ziyade, % 0-% 100 aralığ ında belirtilmektedir (Han ve Kamber, 2000: 243-254). Bu değ erler 1’e ne kadar yakınlarsa nesnel er arasında o kadar güçlü iliş ki olduğ unu göstermektedirler. Dolayısıyla, bağ ıntının önemli olması için her iki değ erin de olabildiğ ince yüksek olması gerekmektedir.

Birliktelik kurallarında başv urulan bir diğ er ölçüt ise kaldıraç (lift) değ eridir. Hesaplaması Tablo 2’deki gibi yapılmaktadır. Kaldıraç değ eri kuralın performansını ve güvenilirliğini ölçümlememizi sağ lamaktadır.

$$Kaldıraç (X, Y) = \frac{P(X, Y)}{P(X)P(Y)} \quad (3)$$

Denklem 3'ün sonucu 1'den küçükse, X'in oluşumu Y'nin oluşumu ile negatif yönde ilişkilidir. Eğer sonuç 1'den büyükse, X ve Y pozitif yönde ilişkilidir. Bu demek oluyor ki, birinin oluşumu diğerinin oluşumunu etkilemektedir. Eğer sonuç 1'e eşitse, X ve Y birbirinden bağımsızdır ve aralarında herhangi bir korelasyon bulunmamaktadır şeklinde yorumlanmaktadır. Bir başka deyişle, olaylardan bir tanesinin meydana gelmesi, diğerinin oluşumunu kuvvetlendirerek, kaldıraç görevi görmektedir. Birliktelik kurallarının elde edilmesi sırasında başvurulabilecek olan birçok algoritma mevcuttur. Bunlardan en sıklıkla başvurulana apriori algoritmasıdır. *Agrawal ve Srikant* tarafından 1994 yılında geliştirilmiştir. Algoritmanın ismi, yaygın nesnelere önsel bilgilerini kullanmasından yani bilgileri bir önceki adımdan almasından “önceki (prior)” kelimesinden gelmektedir. Bu algoritmada sık geçen öge kümelerini bulmak için birçok kez veri tabanını taramak gerekmektedir.

Algoritmanın en bilinen özelliği bir öge kümesi sık geçen ise, bu öge kümesine ait tüm alt kümelerde sık geçen öge kümesi olmalıdır şeklindedir. İteratif bir özelliğe sahip olan Apriori, veri tabanlarında sık geçen öge kümelerinin ortaya çıkarılmasında kullanılmaktadır (Han ve Kamber, 2006: 243-254).

#### 4. Bulgular

Uygulamada veri setine, veri madenciliği birliktelik yöntemlerinden apriori algoritması uygulanmış ve COVID-19 vakalarında hangi ikili, üçlü birlikteliklerin beraber görüldüğü ortaya konmuştur. Kurallarda daha anlamlı birliktelikler elde edebilmek adına, veri setinde şartlı bölümlendirilme yapılmış ve veri seti adım adım incelenmiştir.

Öncelikle sonuçlar cinsiyete göre ele alınmış ve erkek/kadın hastaların hangi şartlar altında (semptom, demografik bilgiler vd.), hastalığı geçirdikleri iyileşme/ölüm durumlarının tespiti yapılmıştır. Kurallar yaş grubuna göre ele alındığında, 35 yaş grubundaki erkeklerde vakaların %89 olasılıkla iyileşme ile sonlandığı tespit edilmiştir. 45 yaş grubundaki erkeklerde ise, vakaların %92 olasılıkla kurtarıldığı tespit edilmiştir. Bu ikili birlikteliğin tüm vakaların %2'sini oluşturduğu tespit edilmiştir.

Kadın hastalar incelendiğinde, ateş semptomunu taşıyan kadın hastaların gözlemlerin %6'sını oluşturduğu tespit edilmiştir. Hastalığı ateşli seyreden kadın hastaların %75 olasılıkla kurtarıldığı tespit edilmiştir. Vakalar semptomlar açısından ele alındığında, elde edilen bulgular şu şekilde olmuştur. Ateş gözlemlenen vakaların %1'lik güven değeri ile ölüm ile sonuçlandığı gözlemlenmiştir. Bu iki birlikteliğin birlikte görülme sıklığı %6 olarak tespit edilmiştir. 75 yaş grubunda ateş gözlemlenen vakalarda, %100 olasılıkla vefat gözlemlenmiştir. Bu birliktelik tüm gözlemlerin %2'sini oluşturmaktadır. Erkek hastaların %65'inde ateş gözlemlendiği ve veri setinde %4'lük bir dilimi oluşturdukları tespit edilmiştir. Kadın hastaların %34'ünde ateş gözlemlendiği ve veri setinde %2'lik bir dilim oluşturdukları tespit edilmiştir. Sadece ateş semptomunu gösterip diğer semptomları (öksürük, boğaz ağrısı, halsizlik) taşımayan vakalar gözlemlerin %7'sini oluşturmaktadır.

Veri setinde yer alan vakalarda Wuhan' dan olan kişilerin %17'sinde ateş gözlemlendiği ve bu kişilerin %1'lik dilimi oluşturdukları tespit edilmiştir. Wuhan' da sadece ziyaret amacı ile bulunmuş kişilerde ise %16'lık oranla ateş gözlemlendiği ve bu hastaların tüm vakaların %1'ini oluşturdukları tespit edilmiştir. 75 yaş grubundaki kişilerin %8'inde ateş gözlemlendiği, 60 yaş grubunda bu oranın %16 olduğu, 50 yaş grubunda ise ateş görülme oranının %13 olarak seyrettiği tespit edilmiştir. Ateş semptomuna ait kuralların bazılarını Tablo 3 ve Tablo 4' de yerverilmiştir.

**Tablo 3.** Ateş semptomuna göre elde edilen kurallar.

	lhs	rhs	support	confidence	lift	count
[1]	{symptom=fever}	=> {death=0}	0.067281106	1.000000000	1.0616438	73
[2]	{symptom=fever}	=> {If_onset_approximated=0}	0.060829493	0.90410959	1.8301472	66
[3]	{symptom=fever}	=> {visiting_wuhan=0}	0.057142857	0.84931507	1.0319226	62
[4]	{symptom=fever}	=> {from_wuhan=0}	0.055299539	0.82191781	0.9640874	60
[5]	{symptom=fever}	=> {recovered=0}	0.049769585	0.73972603	0.8667416	54
[6]	{symptom=fever}	=> {gender=male}	0.044239631	0.65753425	1.3719705	48
[7]	{symptom=fever}	=> {symptom_detailed=}	0.044239631	0.65753425	0.9474431	48
[8]	{symptom=fever}	=> {country=Japan}	0.039631336	0.58904110	3.3637347	43
[9]	{symptom=fever}	=> {i..case_in_country=[50,1.44e+03]}	0.032258065	0.47945205	1.7340183	35
[10]	{symptom=fever}	=> {gender=female}	0.023041475	0.34246575	0.9727103	25
[11]	{symptom=fever}	=> {i..case_in_country=[1,15]}	0.019354839	0.28767123	1.0951694	21
[12]	{symptom=fever}	=> {i..case_in_country=[15,50]}	0.014746544	0.21917808	0.7848456	16
[13]	{symptom=fever}	=> {from_wuhan=1}	0.011981567	0.17808219	1.2385845	13
[14]	{symptom=fever}	=> {age=65}	0.011059908	0.16438356	3.4299262	12

**Tablo 4.** Ateş semptomuna göre elde edilen kurallar (devam).

	lhs	rhs	support	confidence	lift	count
[1]	{age=75,symptom_1=fever}	=> {If_onset_approximated=0}	0.02119816	1	2.024254	23
[2]	{age=75,symptom_1=fever}	=> {visiting_wuhan=0}	0.02119816	1	1.215006	23
[3]	{age=75,symptom_1=fever}	=> {from_wuhan=0}	0.02119816	1	1.172973	23
[4]	{age=75,symptom_1=fever}	=> {death=0}	0.02119816	1	1.061644	23
[5]	{location=Hokkaido,symptom_1=fever}	=> {country=Japan}	0.02488479	1	5.710526	27
[6]	{location=Hokkaido,symptom_1=fever}	=> {If_onset_approximated=0}	0.02488479	1	2.024254	27
[7]	{location=Hokkaido,symptom_1=fever}	=> {death=0}	0.02488479	1	1.061644	27
[8]	{age=65,symptom_1=fever}	=> {visiting_wuhan=0}	0.02396313	1	1.215006	26
[9]	{age=65,symptom_1=fever}	=> {from_wuhan=0}	0.02396313	1	1.172973	26
[10]	{age=65,symptom_1=fever}	=> {recovered=0}	0.02396313	1	1.171706	26
[11]	{age=65,symptom_1=fever}	=> {death=0}	0.02396313	1	1.061644	26
[12]	{age=55,symptom_1=fever}	=> {death=0}	0.02949309	1	1.061644	32
[13]	{location=Hong Kong,symptom_1=fever}	=> {country=Hong Kong}	0.03133641	1	11.542553	34
[14]	{location=Hong Kong,symptom_1=fever}	=> {death=0}	0.03133641	1	1.061644	34

Hastalığın seyrinde görülen öksürük semptomuna ait kurallardan bazılarını Tablo 5’de yer verilmiştir.

**Tablo 5.** Öksürük semptomuna göre elde edilen kurallar.

	lhs	rhs	support	confidence	lift	count
[1]	{symptom_2=cough, source=Government HK}	=> {country=Hong Kong}	0.02764977	1	11.542553	30
[2]	{symptom_2=cough, source=Government HK}	=> {visiting_wuhan=0}	0.02764977	1	1.215006	30
[3]	{symptom_2=cough, source=Government HK}	=> {from_wuhan=0}	0.02764977	1	1.172973	30
[4]	{symptom_2=cough, source=Government HK}	=> {recovered=0}	0.02764977	1	1.171706	30
[5]	{symptom_2=cough, source=Government HK}	=> {death=0}	0.02764977	1	1.061644	30

Veri setinde yer alan lokasyon bilgisine ait detaylı bir inceleme yapıldığında aşağıdaki bulgulara rastlanmıştır. Hong Kong’da hastalığın tespit edildiği kişilerde, sadece öksürük semptomu gözleniyor ise %100

oranla iyileşme olduğu görülmüştür. Bu vakalar toplam gözlemlerin %2'sini oluşturmaktadır. Japonya' da tespit edilen sadece öksürük semptomunu taşıyan vakalarda %100 oranla ölüm gerçekleşmemiştir. Bu vakalar toplam gözlemlerin%2'sini oluşturmaktadır.

## 5. Tartışma ve Sonuç

COVID-19 salgını ile ilgili araştırmalar halen sürmektedir. Salgının ortaya çıkış noktasında kronik rahatsızlıkların ve yaş faktörünün büyük ölçüde etken olduğu bilinmektedir. Elde edilen kurallarda, veri setinde yer alan, ateş, öksürük, boğaz ağrısı, kas ağrısı gibi semptomlar içerisinde en belirgin olarak ateş semptomu ortaya konmuştur. Ateş semptomunu takiben öksürük semptomunun hastalar arasında sıklıkla rastlandığı görülmektedir. Yaş faktöründe özellikle 65 yaş üzerinde semptomların daha hissedilir olduğu tespit edilmiştir. Semptomları taşıyan erkek bireylere kadınlardan daha sık rastlandığı tespit edilmiştir.

Veri setinde yer alan kişiler, Çin'in Wuhan Eyaletinde yaşayan veya bu lokasyonda bulunmuş farklı ülkelerin vatandaşlarından oluşmaktadır. Sonuçlara bakıldığında, Wuhan'da yaşayan kişilerde vakaların ölümle sonuçlanması, ziyaretçi olanlara oranla daha yüksek olduğu görülmüştür. Bu sonuç, virüse maruz kalma sürecinin arttıkça virüsün vücutta tutunma olasılığının artacak olması gerçeği ile örtüşmektedir. Virüsün başlangıç noktası olan Wuhan eyaletinden toplanmış verilerle gerçekleştirilen bu çalışmada hastalığa yakalanan kişilerdeki belirtiler, temel demografik özellikler tespit edilmeye ve ortaya çıkartılmaya çalışılmıştır. Ortaya konulan vaka profilinde erkek bireylerin kadınlara oranla virüse yakalanma oranının yüksek olduğu ve iyileşme oranlarının kadınlara göre daha düşük olduğu tespit edilmiştir. Dünyayı etkisi altına alan bu salgın ile ilgili çalışmalar, artan veriler doğrultusunda gelişmeye ve devam edecektir. Bu çalışmanın da temin edilen yeni veriler ışığında dönemsel olarak tekrarlanabileceği ve böylelikle

güncelliğini koruyarak elde edilen bulguların sağlık bilimine faydalı olacağı düşünülmektedir.

## Kaynakça

- Agrawal, R. ve Srikant, R. (1994). Fast Algorithms for Mining Association Rules, *Proceedings of the 20th International Conference on Very Large Databases*, ss. 487–499, 1994.
- Alanlar, E. (2021). *Pazar Sepeti Analizi ile Birliktelik Kurallarının Belirlenmesi: Perakende Sektöründe Covid-19 Etkisi*, Yüksek Lisans Tezi, Karabük Üniversitesi, Eğitim Enstitüsü.
- Atsa'am, D. D., ve Wario, R. (2022). Association Rules on the COVID-19 Variants of Concern to Guide Choices of Tourism Destinations, *Current Issues in Tourism* (Vol. 25, Issue 10, 1536–1540). <https://doi.org/10.1080/13683500.2021.1951182>
- Birant, D., Kut, A., Ventura, M., Altınok, H., Altınok, B., Altınok, B. ve Ihlamur, M. (2010) İş Zekası Çözümleri için Çok Boyutlu Birliktelik Kuralları Analizi, ss. 215–222.
- Çelik, A. (2020). Using Apriori Data Mining Method in COVID-19 Diagnosis. *Journal of Engineering Technology and Applied Sciences* 5 (3) 2020: 121-131.
- Dong, E., Du, H., ve Gardner, L. (2020). An Interactive Web-Based Dashboard to Track COVID-19 in Real Time, *The Lancet Infectious Diseases*, 20(5), 533–534. [https://doi.org/10.1016/S1473-3099\(20\)30120-1](https://doi.org/10.1016/S1473-3099(20)30120-1)
- Er, A.G. ve Ünal, S. (2020). 2019 Koronavirüs Salgını – Anlık Durum ve İlk İzlenimler. *Flora*, 2020 (25).
- Han, J. ve Kamber, M. (2006). *Data Mining: Concepts and Techniques* (The Morgan Kaufmann Series in Data Management Systems), 2nd Edition, Waltham, USA, Morgan Kaufmann Publishers, 2006.
- Hahsler, M., Buchta, C., Gruen, B. ve Hornik, K. (2022). *arules: Mining Association Rules and Frequent Itemsets*. R package version 1.7-3, <https://CRAN.R-project.org/package=arules>.
- Hahsler, M. (2021). *arulesViz: Visualizing Association Rules and Frequent Itemsets*. R Package Version 1.5-1, <https://CRAN.R-project.org/package=arulesViz>.
- Ji, Y., Ma, Z., Peppelenbosch, M. P., ve Pan, Q. (2020). Potential Association Between COVID-19 Mortality and Health-Care Resource Availability. In *The Lancet Global Health* (Vol. 8, Issue 4, p. e480). Elsevier Ltd. [https://doi.org/10.1016/S2214-109X\(20\)30068-1](https://doi.org/10.1016/S2214-109X(20)30068-1)

- Liu, R., Ming, X., Zhu, H., Song, L., Gao, Z., Gao, L., Jia, S., Zhang, C., Xu, O., Zhou, Yuan, J., ve Zhang, J. (2020). “Association of Cardiovascular Manifestations with In-Hospital Outcomes in Patients with COVID-19: A Hospital Staff Data”. *MedRxiv*, 1-26,2020.
- Nahar, J., Imam, T., Tickle, K. S., Chen, Y.-P. P., Yang, H. H., Fong, S., Lara, J. a., Lizcano, D., Martínez, M. a., Pazos, J., Riera, T., Yang, H. H., Chen, Y.-P. P., Nisbet, R., Elder, J., Miner, G., Garg, S., Sharma, A., Han, J., Kamber, M., Contributions, O., Koh, H. C., Tan, G., Narwal, S. ve Zhao, Y. (2013) ‘Association Rule Mining to Detect Factors which Contribute to Heart Disease in Males and Females’, *Expert Systems with Applications*. Elsevier Ltd, 102(3), 335–351. doi: 10.1016/j.jss.2014.07.010.
- Shearer, C. (2000). JOURNAL Statement of Purpose E-Business and the New Demands on Data E-Commerce Places on Data Warehousing Technology. *Journal of Data Warehousing*, 5(4), 13–22.
- Tandan, M., Acharya, Y., Pokharel, S., ve Timilsina, M. (2021). Discovering Symptom Patterns of COVID-19 Patients Using Association Rule Mining. *Computers in Biology and Medicine*, 131. <https://doi.org/10.1016/j.compbimed.2021.104249>
- URL-1: <https://www.kaggle.com/datasets/sudalairajkumar/novel-corona-virus-2019-dataset>, (Erişim Tarihi: 05.01.2022).
- URL-2: R-project, 2022, *The R Project for Statistical Computing*, <http://www.r-project.org/>, (Erişim Tarihi: 28.02.2022).
- URL-3: RStudio, 2022, *Home - RStudio*, <http://www.rstudio.com/>, (Erişim Tarihi: 28.02.2022).
- WHO. “Coronavirus Disease 2019 (COVID-19) Situation Reports”. <https://www.who.int/emergencies/diseases/novelcoronavirus-2019/situation-reports> (04.07.2022).
- WORLDOMETERS. (2022). Covid-19 Coronavirus Pandemic. <https://www.worldometers.info/coronavirus/> (25.07.2022).
- Wojtusiak, J., Bagchi, P., Durbha, S.S.K.R.T.N. et al. COVID-19 Symptom Monitoring and Social Distancing in a University Population. *J Healthc Inform Res* 5, 114–131 (2021). <https://doi.org/10.1007/s41666-020-00089-x>
- Wu, Z., & McGoogan, J. M. Characteristics of and Important Lessons from the Coronavirus Disease 2019 (COVID-19) Outbreak in China: Summary of a Report of 72314 Cases from the Chinese Center for Disease Control and Prevention. *JAMA - Journal of the American Medical Association*, 2019, 3–6, 2020.
- Xia, S., Zhong, Z. ve Gao, B. (2021). The Important Herbal Pair for the Treatment of COVID-19 and its Possible Mechanisms. *Chin Med* 16, 25. <https://doi.org/10.1186/s13020-021-00427-0>



- Xu, B., Gutierrez, B., Mekaru, S., Sewalk, K., Goodwin, L., Loskill, A., Cohn, E. L., Hswen, Y., Hill, S. C., Cobo, M. M., Zarebski, A. E., Li, S., Wu, C. H., Hulland, E., Morgan, J. D., Wang, L., O'Brien, K., Scarpino, S. V. V., Brownstein, J. S. ve Kraemer, M. U. G. (2020). Epidemiological Data From the COVID-19 Outbreak, Real-Time Case Information. *Scientific Data*, 7(1), 1–6. <https://doi.org/10.1038/s41597-020-0448-0>
- Zhang, C. ve Zhang, S. (2007) *Association Rule Mining Models and Algorithms*, Berlin, Springer, ISBN: 3-540-4353.

