aurum

AURUM MÜHENDİSLİK SİSTEMLERİ VE MİMARLIK DERGİSİ
AURUM JOURNAL OF ENGINEERING SYSTEMS AND ARCHITECTURE

Cilt 3, Sayı 2 | Kış 2020
Volume 3, No 2 | Winter 2020, 153-168

ARAŞTIRMA MAKALESİ / RESEARCH ARTICLE

VERİ MADENCİLİĞİ YÖNTEMLERİ İLE UÇUŞ BİLETLEME ANALİZİ*[1]

Muhammed Metin ULUYARDIMCI[1]
[1]İstanbul Aydın Üniversitesi Fen Bilimleri Enstitüsü Bilgisayar Mühendisliği Yüksek Lisans Programı, İstanbul, Türkiye
muluyardimci@gmail.com ORCID: 0000-0002-6671-2874

Metin ZONTUL[2]
[2]İstanbul Arel Üniversitesi, Mühendislik ve Mimarlık Fakültesi, İstanbul, Türkiye
metinzontul@arel.edu.tr ORCID: 0000-0002-7557-2981

GELİŞ TARİHİ / RECEIVED DATE: 20.02.2019 KABUL TARİHİ / ACCEPTED DATE: 09.04.2019

Özet

Bir çok sektörde kullanılan veri madenciliği hava yolu şirketleri açısında da büyük potansiyel barındırmaktadır. Kritik öneme sahip müşteriye direkt dokunan stratejik kararlarda, veri madenciliği yöntemleri etkin kullanılmaktadır. Verinin işlenip bilgiye dönüştürülme sürecine veri madenciliği denir. Birliktelik kuralları ve Apriori algoritması veri madenciliği alanında sıklıkla kullanılan yöntemlerdir.
Bu tez çalışmasında, öncelikle veri madenciliği açıklanmış ve çalışmada kullanılacak yöntemler tanıtıldıktan sonra Türk Hava Yollarının 2016 yılı yaz ve kış dönemlerine ait yolcuların biletleme verileri ele alınmıştır. Veri ön işleme ve temizleme süreçlerinden sonra 2036113 satırdan oluşan uçuş biletleme verisine Birliktelik kuralları ve Apriori Algoritması uygulanarak 824 adet kural ortaya çıkarılmıştır. Elde edilen kurallar yorumlanarak ve kuralların etkileri değerlendirilerek müşteri deneyimine olumlu katkı sağlayabilecek öneriler sunulmuştur.

Anahtar Kelimeler : Veri Madenciliği, Apriori Algoritması, Birliktelik Kuralları, Uçuş Biletleme

FLIGHT TICKETING ANALYSIS WITH DATA MINING METHODS

Abstract

Data mining used in many sectors also has great potential in terms of airline companies. Data mining methods are used effectively in strategic decisions that directly touch the critical customer. The process of processing the data into information is called data mining. Association rules and Apriori algorithm are frequently used methods in the field of data mining.
In this thesis, firstly data mining was explained and after the introduction of the methods to be used in the study, the ticketing data of the passengers of the 2016 summer and winter periods of Turkish Airlines were discussed. After pre-processing and cleaning processes, 824 rules were applied to the flight ticketing data consisting of 2036113 lines by using Association Rules and Apriori Algorithm. By interpreting the obtained rules and evaluating the effects of the rules, suggestions are presented that can contribute positively to the customer experience.

Keywords: Data Mining, Apriori Algorithm, Association Rules, Flight Ticketing

1 * Bu makale büyük ölçüde "Veri Madenciliği Yöntemleri İle Uçuş Biletleme Analizi, İstanbul Aydın Üniversitesi Fen Bilimleri Enstitüsü Bilgisayar Mühendisliği Yüksek Lisans Tezi, 2018" den yararlanarak hazırlanmıştır.

1. GİRİŞ

Teknolojideki hızlı ve büyük gelişimler her geçen gün daha da ilerleme kaydetmektedir. Bu gelişimlerden bilgi sistemleride olumlu şekilde etkilenmiştir. Bilgi sistemlerinin teknik kapasiteleri artarken ters orantılı olarak maliyetlerinin azalması olumlu etkilerinden sadece biridir. Veri depolama araçlarının fiziksel yapılarının küçülmesi ama teknik kapasitelerinin aynı oranda büyümesi daha çok veri daha az maliyet sağlaması kurum ve kuruluşların teknolojiye olan ilgilerini artırmış ve teknoloji daha fazla kullanmaya yöneltmişdir. Böylelikle, kurum ve kuruluşlar artık daha fazla sayısal olarak veri toplama ve veriyi depolama imkan bulmuştur.

Bu teknolojik gelişmeler günlük hayatımızda kullandığımız bir çok araç ve gereçleri de doğrudan etkilemiştir. Hayatımıza, birbirimiz ile görüşmek mesaj atmak gibi iletişim için giren cep telefonları artık iletişimin yanı sıra geliştirilen mobile uygulamalar ile bir çok sektördeki bir çok işlemi gerçekleştirdiğimiz, iş takibi yaptığımız bir araç halini almıştır. Artık alışveriş, eglence, ulaşım gibi bir çok ihtiyac, bu teknolojiler kullanılarak gerçekleştirilebiliyor. Kullanılan bu teknolojik imkanlar yardımıyla yapılan tüm işlemlerin detayları sayısal veri olarak saklanabilmektedir. Bu sayede kullanıcıların ayrıntılı hareket bilgileri takip edilerek kullanıcıların daha fazla alışveriş yapması ve alışveriş esnasında farklı nesnelere yönelmeleri için çeşitli kampanyalar veya etkinlikler düzenlenmektedir.

Saf ve işlenmemiş veri, kendi başına bir anlam ifade etmeyebilir. Belirli bir amaç doğrultusunda üzerinde yapılacak bir takım işlemler sonrasında bilgi elde edilebilir. Veriden bilgi'ye erişilmesine veri analizi denir (Akpınar, 2000).

Bir çok alanda kullanıldığı gibi veri analizi ulaştırma sektörlerinden biri olan hava yollarında da önemli bir yer almaktadır. Hava yolu müşterilerinin yapmış oldukları tüm işlemler detaylı bir şekilde sayısal olarak veri tabanlarında tutulmaktadır. Sürekli artan bu verilerin incelenmesi, analiz edilmesi için de yazılımlara ihtiyaç duyulmaktadır.

Bu aşamada da veri madenciliği tekniklerine ihtiyaç duyulur. Veri madenciliği tekniklerinden biri olan Apriori algoritması verideki nesneler arası ilişkileri ortaya çıkarmak için kullanılır (Eker, 2016). Bu veriler müşterilerin yapmış olduğu uçuşlarda tercih ettikleri bir çok faktörün göz önüne alınarak gelecekti yatırımların ya da çalışmaların yapılmasında büyük bir rol oynamaktadır.

Bu çalışmanın amacı, Türk Hava Yolu şirketi müşterilerinin yıl içerisinde dönemsel olarak yapmış olduğu uçuş biletleme işlemlerinin birleştirme kuralları ve apriori algoritması kullanılarak yapılan analiz sonrası ortaya çıkan sonuçların yorumlanmasıdır. Yapılan benzer çalışmalar ve bu çalışmanın literatüre katkısı aşağıda verilmiştir.

1.1. Literatür Taraması

Bilgi sistemlerinin ve teknolojilerinin gelişmesi, verinin değerini gittikçe daha önemli bir hale getirmiştir. Verinin hızlı bir şekilde işlenebildiği ve gelişmiş veri analizi yapılabilmesine bağlı olarak veri analizinden daha anlamlı sonuçlar elde edilebilmesi (Sönmez, 2018) ile, bir çok kurum ve kuruluş kendi amaçları doğrultusuda veritabanlarında bir çok türde veri depolamıştır. Ulaştırma, pazarlama işlemleri, kamusal

alandaki işlemler, biletleme işlemleri ve buna benzer bir çok alanda saklanan büyük boyuttaki ve karmaşık verilerden anlamlı kuralların ortaya çıkarılmasına ihtiyaç duyulmaktadır. Keşfedilmemiş ve işlenmemiş bu verilerden yeni, geçerli, faydalı ve sonuç olarak anlaşılabilir örüntülerin çıkarılmasındaki bu bilgi keşfi sürecine Veritabanlarında Bilgi Keşfi (Knowledge Discovery in Databases - KDD) denir (Döşlü, 2008).

Veritabanlarında bilgi keşfi sürecinin bir aşaması olarak bilinen Veri Madenciliği (Data Mining), geçerli, yeni ve kullanışlı bilgiyi büyük veri tabanlarından çıkarma işlemidir. Bu yüzden veri madenciliği, veri tabanından anlamlı örüntüler veya kurallar elde etmek için geniş bir araştırma alanı olarak görülmektedir Veri madenciliği, veri tabanlarında bilgi keşfi sürecinde anlamlı örüntüleri elde eden keşif algoritmaları ile veri analizini uygulayan bir adımdır. Veri madenciliği, veri tabanlarındaki bilgi keşfi uygulamaları ile birlikte faaliyet alanına yönelik karar destek mekanizmaları için gerekli ön bilgileri temin etmek için kullanılmaktadır (Fayyad ve Ark., 1996).

"Bir başka ifadeyle büyük miktardaki verinin analiz edilerek anlamlı şablon ve kuralların keşfedilmesine imkan verir" (Berry ve Linoff, 2004).

Veri madenciliği 1990'lı yıllarda ortaya çıkmıştır. Veri madenciliğinin özellikle 2000 yılından bu yana büyük bir gelişme gösterdiği göze çarpmaktadır (Gülce, 2010).

Aşağıda veri madenciliği ve Apriori algoritması konusunda farklı alanlarda yapılan bilimsel çalışmalar incelenmiştir.

Yapılan bir makale çalışmasında Çin'deki ciddi trafik kazalarına yol açan faktörleri araştırmak amaçlanmıştır. 2009 ile 2013 yılları arasında çin'de meydana gelen trafik kazalarının tutulduğu faaliyet raporundaki veriler alınarak veri madenciliği yöntemlerinden biri olan birliktelik kuralı uygulanmış ve sonuç olarak ciddi trafik kazalarının kullanıcı davranışı, yolun geometrik özellikleri ve çevresel faktörler arasındaki karmaşık etkileşimlerin bir sonucu olduğunu ortaya çıkarmıştır (Xu ve Arkadaşları, 2018).

Muhammed Emin Eker 2016 yılında hazırladığı yüksek lisans çalışmasında, bir eğitim yazılımındaki verilerden bilginin ortaya çıkarılmasını amaçlamıştır. Erişilen verilere veri madenciliğinde en sık kullanılan apriori algoritması ve birliktelik kuralları ayrıntılı bir şekilde işlenmiştir. Uygulama esnasında hazırlanan yazılım, verilerin elde edildiği eğitim yazılımına dâhil edilmiştir. Bu sayede yapılan her sınav sonrasında oluşan sınav verileri içerisinden anında ilişki kuralları çıkarılabilmektedir. Yapılan çıkarımlar, bu eğitim yazılımını kullanan kurum ve kişilerin hizmetine sunulduğu belirtilmiştir (Eker, 2016).

Mehmet Aydın Ulaş 1999 yılında alışveriş sektöründe (alanında) hazırladığı yüksek lisans tez çalışmasında, sepet analizi gerçekleştirilmiştir. Süpermarket zinciri olan Gima Türk A.Ş.'nin verileri üzerine Apriori algoritması uygulanmış ve ulaşılan sonuçlar incelenmiştir. Ayrıca mal satışları arasındaki ilişkileri bulmak amacıyla da, bileşen analizi ve k-means metotları kullanılmıştır (Ulaş, 1999).

Yapılan bir başka çalışmada, birliktelik kuralları için bir yöntem önermiştir. Bir elektronik firmasında üretim ve mal giriş kalite verileri üzerinde Apriori algoritmasının oluşturduğu kurallar elenerek uygulanmıştır. Elde edilen kurallar test verileri ile doğrulanmış ve sonuçlar analiz edilmiştir (Kılınç, 2009).

Döviz piyasalarındaki uluslar arası para birimleri arasındaki iç ilişkilerin ele alındığı bu makale çalışmasında, 2011 ile 2016 yıllarına ait dönemlerde Tayvan yatırımcılarının döviz portföylerinin tanımlamak ve değerlendirmek amaçlanmıştır. Tayvan doları ile 15 ülkenin döviz kuru verilerine birliktelik kuralı ve apriori algoritması uygulanarak ortaya çıkartılan sonuçlar değerlendirilmiştir (Lai ve Jin, 2018).

Barış Yıldız 2010 yılında hazırladığı yüksek lisans çalışmasında, sık kümelerin ortaya çıkarılması için gizliliği koruyan bir yaklaşım sunmuştur. Bu çalışmayla beraber ayrıca, Matrix Apriori algoritması üzerinde değişiklikler yapılmış ve sık küme gizleme çerçevesi de geliştirilmiştir (Yıldız, 2010).

2. VERİ MADENCİLİĞİ

Veri madenciliği, günümüz bilgi çağında en güncel teknolojilerden birisidir. Bilgisayar sistemlerinin her geçen gün daha da gelişmesi ve güçlerinin artıyor olması, veri tabanlarında daha büyük miktarlarda verinin saklanabilmesine imkan vermektedir. Veri madenciliği, veri analizi ile gelişmiş matematiksel algoritmalar kullanılarak, kalıpları ve eğilimleri keşfetderek gelecekteki olayların olasılığını değerlendirmek için büyük veri kümeleri arasındaki sıralama sürecidir.

Veri madenciliği ile ilgili literatürde farklı araştırmacılar tarafından yapılan tanımlamalar aşağıdaki gibidir.

"Veri madenciliği büyük miktarda veri içinden gelecekle ilgili tahmin yapmamızı sağlayacak bağıntı ve kuralların bilgisayar programları kullanarak aranmasıdır" (Akpınar, 2000).

"Veri madenciliği, büyük veri setindeki, anlamlı, orijinalliği olan, kullanım potansiyeli bulunan ve sonuçta anlaşılabilir olan örüntülerin çıkarılmasıdır" (Fayyad ve Ark.,1996).

"Veri madenciliği, veri içeriğinde yapılan uygulamalar sonrasında veriler arasında bağlantının kurulması amaclı bir algoritma çalıştırma işlemidir" (Zaimoğlu, 2018).

"Tek başına ham verinin sunamadığı bilgiyi ortaya çıkaran veri analizi sürecine veri madenciliği denir" (Jacobs, 1999).

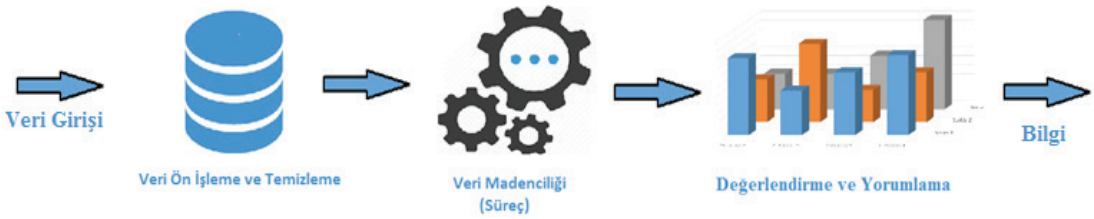
"Veri madenciliği, büyük ölçekli veriler arasından değeri olan bir bilgiyi elde etme işidir" (Özkan, 2008).

2.1. Veri Madenciliği ve Bilgi Keşfi

Veri madenciliği, veri tabanında bilgi keşfi sürecinde temel bir adımdır (KDD). Şekil 2.1'de gösterildiği gibi genel (KDD) süreci, üç ana adımdan oluşur.

aşağıdaki gibidir:

- Veri ön işleme ve temizleme
- Veri madenciliği
- Veri sonrası işlem (Değerlendirme ve Yorumlama)



Şekil 2.1: Veri Madenciliği Süreci (Özkan, 2008)

Veri ön işleme ve temizleme, ham veri toplamak ve hazırlamak için kullanılan bir süreçtir.

Veri ön işleme ve temizleme dört aşamadan oluşmaktadır.

Veri ön işleme ve temizleme aşamaları şunlardır:

- Veri temizleme
- Veri entegrasyonu
- Veri seçimi
- Veri dönüşümü

Veri temizleme: Çeşitli kaynaklardan veya mevcut bir sistemden temin edilen verilerden anlamlı bir bilgi ortaya çıkarabilmek için kirli veya kayıp verilerin tespit edilerek çeşitli işlemler uygulanmasıdır (Zaimoğlu, 2018).

Veri entegrasyonu: Farklı kaynaklardaki verilerin birlikte derlenerek tek bir veri tipine dönüştürülme işlemidir.

Veri seçimi: Veri öbeklerinin analiz aşamasında sonucu etkilemediği tespit edilen veri sayısı veya değişkenin çıkarılma işlemidir. Böylelikle tespit edilen gereksiz verilerin analizden çıkarılması ile boyut azalması yapılır (Özkan, 2008).

Veri dönüşümü: Algoritmanın çalışmasına anlam bakımından uygun olmayan veri içeriğinin belirlenen bir işlem ile dönüştürülerek kullanıma alınabilir bir duruma getirilmesidir. Bu tür verilerde normalleştirme ya da standartlaştırma gibi süreçler ile veri dönüşümü sağlanır (Zaimoğlu, 2018).

Veriler hazırlandıktan sonra yüksek seviyeli bilgileri ayıklamak için kurallar ve yöntemler uygulanılarak yararlı bilgiler ayrıştırılır ve bu süreç veri madencliği süreci olarak adlandırılır. Elde edilen bilgiler görselleştirme ve diğer teknikler ile sunulabilmektedir (Al-Rubaiee, 2018).

Veri sonrası işlemler'de, ham olarak bilinen bilgi kendi başına değersizdir ve gerçek bilgi değildir. Bilgi, verileri ile desteklendiği ve yorumlandığı zaman gerçekler haline gelir. Bu aşamada bilgi, bilginin iletişimdir.

2.2. Veri Madenciliğinin Gelişimi

1960'larda bilgisayarların veri analizi amacıylada kullanılmaya başlanmasıyla birlikte veri madenciliği kavramsal olarak ortaya çıkarmıştır. O yıllarda, yeterince uzun taramalar yapılması durumunda istenilen verilere ulaşmanın mümkün olacağına inanılıyordu. Yapılan bu işleme veri madenciliği yerine daha önceleri veri taraması (data dredging), veri yakalanması (data fishing) gibi adlandırıldığı bilimektedir (Öğüt, 2005).

1980'lerde bağıntılı (relational) veritabanları ve SQL (Select Query Language) yapısal sorgulama dili ile verilerin dinamik ve anlık analiz edilmesine olanak sağlanmıştır (Altun, 2017).

1990'lı yıllara gelindiğinde toplanan verilerin hacmi çok büyük boyutlara ulaşmış ve verilerin depolanması için veri ambarları kullanılmaya başlanmıştır (Altun, 2017). Yine 1990'larda veri madenciliğine farklı yaklaşımlar getirilmeye başlanmıştır.

Bu yaklaşımların kökeninde pazarlama, otomasyon, istatistik, veritabanlar ve makine öğrenimi gibi disiplinler ve kavramlar bulunmaktaydı (Öğüt, 2005).

2000'li yıllar veri madenciliğinin en yaygın olduğu ve tüm alanlarda veri madenciliğinin kullanıldığı yıllar olarak bilinmektedir. Veri madenciliği, depolanan bu büyük veri kütlelerinin değerlendirilmesi için yapay zekâ ve istatistik tekniklerinin uygulanması sonucunda ortaya çıkmıştır (Altun, 2017).

2.3. Veri Madenciliği Yöntemleri

Veri madenciliği uygulanan verilerde istenilen analizlere ulaşmak için veri setine uygun yöntem kullanmak ve bu yöntemi uygularken veriyi de yönteme uygun hale getirmek verinin doğru işlenmesi açısından önem arz etmektedir (Zaimoğlu, 2018).

Veri madenciliği yöntemleri genel olarak iki ana kategoriye ayrılır

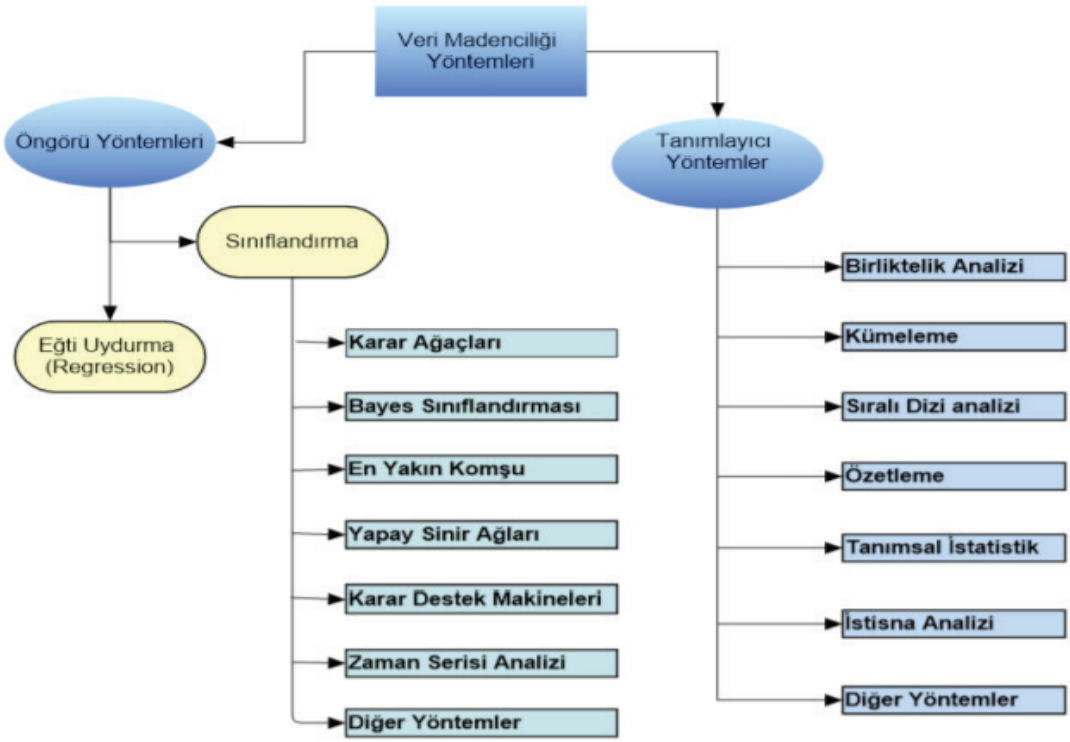
2.3.1. Öngörü yöntemi

Öngörü modelleme çalışması sonucunda belirli bir konuda öngörü için diğer özelliklerin değerlerine dayanarak kullanılacak bir model oluşturmayı amaçlar. Örneğin bir fabrikanın bir önceki yıla ait sipariş verilerini kullanılarak gelecek yıla yönelik üretim planlaması yapılması öngörü modeline bir örnektir.

Sınıflandırma ve regresyon olmak üzere öngörü modeli iki ana kısımdan oluşmaktadır (Zaimoğlu, 2018).

2.3.2. Tanımlayıcı model

Veri tabanındaki analiz edilen verilerin sonucu, verinin mevcut durumunu ya da yaklaşık olarak neyi ifade ettiğini belirten yöntemleri kapsar. Henüz tespit edilmemiş ya da önceden keşfedilmemiş bilgiyi tespit etmek, bulmak için kullanılan bir yöntemdir (Zaimoğlu, 2018). Örneğin veri tabanında amaçlanan bir veri setini bulabilmek için aynı anda bir çok vasfı içeren bilginin durmunu ve veri kümesindeki örüntünün ona eş olanlarını belirlemek tanımlayıcı modele örnektir. Tanımlayıcı model, kümeleme, birliktelik kuralları ve ilişkilendirme kuralları çözümleme en sık kullanılan kurallardır (Al-Rubaiee, 2018). Şekil 2.2'te veri madenciliği tanımlayıcı ve öngörü yöntemleri listelenmiştir.



Şekil 2.2: Veri Madenciliği Tanımlayıcı ve Öngörü Yöntemleri (Zaimoğlu, 2018)

Veri madenciliğinde Öngörü ile tanımlayıcı model yöntemlerinin anlaşmaları halinde birbirlerinin yerlerine kullanılabilirler. Bu modeller kuşkusuz birbirlerinden ayrılamazlar (Zaimoğlu, 2018).

2.4. Veri Madenciliği Uygulama Alanları

Veri madenciliği, günümüz bilgi çağında en güncel teknolojilerden birisidir. Bilgisayar sistemlerinin her geçen gün daha da gelişmesi ve güçlerinin artıyor olması, veri tabanlarında daha büyük miktarlarda verinin saklanabilmesine imkan vermektedir. Bu büyük miktardaki verilerden faydalı bilgilere ulaşılması her sektör için gün geçtikce veri madenciliği tekniklerinin uygulanmasını sağlamıştır.

Aşağıda veri madenciliği tekniklerinin uygulandığı sektörler özetlenmiştir.

- Üretim ve İmalat
- Devlet Uygulamaları
- Bankacılık ve Finans Uygulamaları
- Biyomedikal ve DNA
- Haberleşme ve İletişim
- Mühendislik Uygulamaları

- Pazarlama
- Ulaştırma
- Eğitim
- E-Ticaret
- Sigortacılık
- Sağlık

2.5. Birliktelik Kuralları

Birliktelik kuralı veri madenciliği, ilişkisel veritabanları, işlem veritabanları ve diğer veri havuzu formları gibi farklı veri tabanlarında bulunan veri kümelerinden sık kalıpları, bağıntıları, ilişkileri ya da nedensel yapılara ulaşmak için tutulan yol ve yöntemdir. Veri madenciliğinde birliktelik kuralları, süregelen bir takım işlem göz önüne alındığında işlemdeki diğer öğelerin oluşmalarına dayanarak belirli bir öğenin oluşumunu tahmin etmemizi sağlayan kuralları bulmayı amaçlar. Kısaca verilerin birlikte bulunma durumlarının analiz edilerek tespit edilmesinde birliktelik kuralları kullanılmaktadır (Özkan, 2008).

Birliktelik kurallarında güven ve destek ilgi çekici iki ölçüsüdür. Bunlar kullanıcı tarafından sağlanan parametrelerdir ve kullanıcıdan kullanıcıya değişir (Sadıqmal, 2015).

Birliktelik kuralları uygulamalarında, örnek olarak piyasa sepet analizi verebiliriz. Sepet Analizi, marketin içinde müşterilerin farklı satın alma alışkanlıklarını analiz ederek müşteriler tarafından satın alınan öğeler arasındaki alışkanlıklarını ortaya çıkarılmasıdır (Zaimoğlu, 2018).

Piyasa sepeti analizinde örneğin *"müşteriler bira satın aldığında %75 ihtimalle cipste satın alırlar"* sonucu ulaşarak bu iki ürün arasında güçlü bir ilişki olduğu tespit edilir (Farboudi, 2009).

2.5.1. Güven destek ve diğer kavramları

Kısaca, verilerin analizi sonucu oluşan kurallar arasından bir ya da bir çok yararlı kural kümesini keşfetmekle ilgilidir. Veri madenciliğinde, birliktelik kuralları ile çok sayıda kurala erişilebilinmektedir. Amaç yararlı olan kuralları bulabilmektir. Yararlılığı ölçmenin yolu güven ve destek değerlerinden geçmektedir. Bu değerler niteliğe göre keşfedilen kuralların kullanışlığını ve doğruluğunu ifade eder (Gülce, 2010).

Güven ve destek değerlerinin örnek bir formülü aşağıdaki gibir:

X ürünün yapıldığı satışlarda Y ürününde alınması durumu [destek = % 4, güven = % 55]

X ürünün yapıldığı satışlarda Y ürününde alınması olayının, güveni aşağıdaki gibi hesaplanabilir **(2.1)**:

$$\text{Güven} = \frac{\text{X ve Y'nin bulunduğu satır sayısı}}{\text{X'sın bulunduğu satur sayısı}}$$

(2.1)

Güven oranı %55 olan durumda; X ürünü almak isteyen müşterilerin %55'ı Y ürününüde satın almak istemiştir. Güven değerinin yüzde yüz olması bu iki ürünün aynı anda alınması anlamına gelir ve bu tür kurallar kesin kural olarak adlandırılır (Gülce, 2010).

X ürününün yapıldığı satışlarda Y ürünününde alınması olayına, desteği ise aşağıdaki gibi hesapnabilir **(2.2):**

$$\text{Destek} = \frac{\text{X ve Y'nin bulunduğu satır sayısı}}{\text{Toplam satır sayısı}}$$

(2.2)

Destek oranı %4 olan durumda; Yapılmış olan tüm satışların % 4'sinde X ürünü ve Y ürünü birlikte bulunmaktadır.

TID	ÜRÜNLER
1	A,B,C,D
2	A,C,B,E
3	F,B,C,D
4	B,C,D
5	A,F,C,D
6	A,C,D

Şekil 2.3: Ürünler Satış

Şekil:2.3.'deki tabloya baktığımız zaman ürünlerin satış hareketlerine göre (C, D) ürünlerinin yapıldığı satışlarda A ürününde alınması arasındaki ilişki, aşağıdaki örnekle açıklanabilir.

$$\text{Güven} = \frac{(C, D \rightarrow A)\quad 3}{(C, D)\quad 5} = 0.6$$

(2.3)

$$\text{Destek} = \frac{(C, D \rightarrow A)\quad 3}{\text{Toplam hareket}\quad 6} = 0.5$$

(2.4)

Yapılan güven ve destek değerlendirmeleri sonucu (C, D) ürünlerinin yapıldığı satışlarda A ürününde alınmasında **(2.4)** %50 destek, **(2.3)** %60 güven oranlarına ulaşılmıştır.

Lift, genel olarak A ve C olaylarının istatistiksel olarak bağımsız olmaları durumunda ne kadar sıklıkla gerçekleştiğini ölçerek ortaya çıkarılmasını sağlar. Lift değeri aşağıdaki formülü kullanılarak hesaplanır (Lia ve Lu, 2018).

$$\text{Lift}(A \rightarrow C) = \frac{\text{Confidence}(A \rightarrow C)}{\text{Support}(C)}, \text{aralık: } [0, \infty]$$

(2.5)

Levarage, A ve C' nin birlikte gözlemlenen sıklığı ile A ve C' nin bağımsız olması durumunda gözlemlenen sıklığı arasındaki farkı hesaplar. Levarage değerinin sıfır olması, iki olayın bağımsız olduğunu ortaya çıkartır. Levarage hesaplanır iken aşağıdaki formül kullanılır (Lia ve Lu, 2018).

$$\text{Levarage}(A \rightarrow C) = \text{support}(A \rightarrow C) - \text{support}(A) * \text{support}(C), \text{aralık: } [-1,1]$$

(2.6)

Conviction, A ve C'nin birlikte ya da A ürününün C ürünü olmaksızın görülme olasıkları hesaplanması için kullanılır. Formülü aşağıdaki gibidir (Lia ve Lu, 2018).

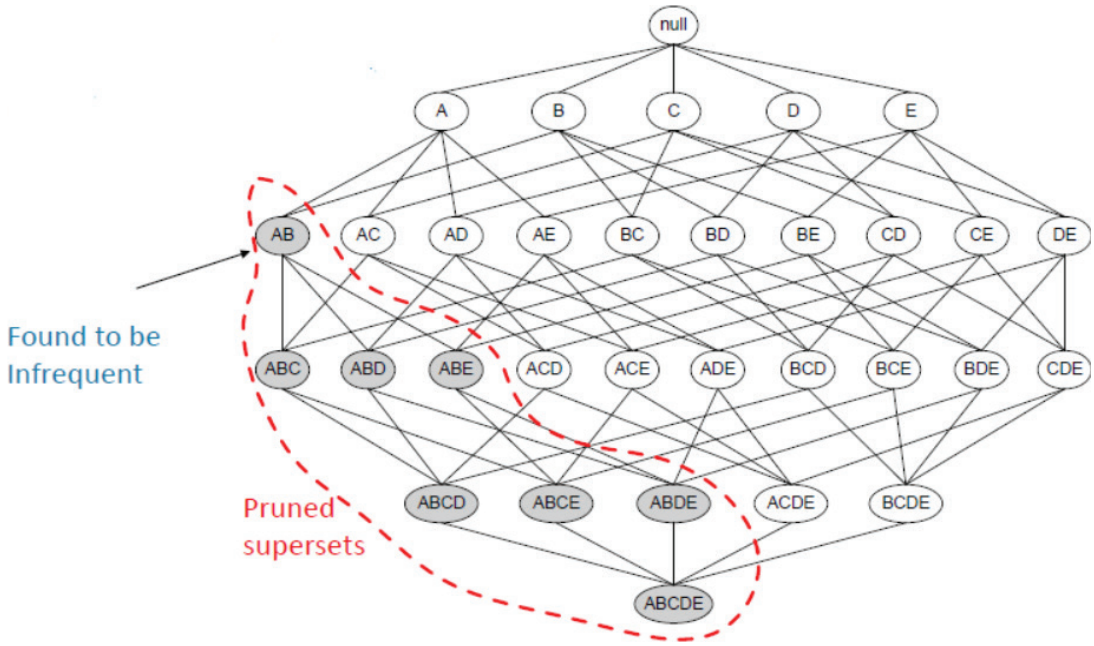
$$\text{Conviction}(A \rightarrow C) = 1 - \frac{\text{Support}(C)}{1 - \text{Confidence}(A \rightarrow C)}, \text{aralık: } [0, \infty]$$

(2.7)

2.6. Apriori Algoritması

Apriori algoritması veri madenciliğinde klasik bir algoritmadır. Birliktelik kuralı uygulanmış verilerdeki ilişki kurallarını tespit etmek için kullanılan en popüler algoritmadır. İlişki analizinde geniş nesne kümelerinin tespit edilerek ortaya çıkarılmasını amaçlamaktadır (Eker, 2016).

Amaçlanan geniş nesne kümelerini ortaya çıkarmak için, ilk olarak her bir nesnenin destek oranı matematiksel bir işlem uygulanarak hesaplanır ve belirtilen destek oranı ise karşılaştırılır. Destek oranı hesaplananlara aday nesne kümesi, belirtilen destek oranını aşan nesne kümesine ise geniş nesne kümesi denilmektedir. Apriori algoritmasının çalışma şekli, bu geniş nesne kümelerinin ortaya çıkararak destek seviyesinin altında kalan nesne kümelerini bir sonraki adımda taramayarak en geniş nesne kümesini tespit edinceye kadar tüm veriler içerisinde tarama yapmaktır (Eker, 2016). Şekil 2.4'de apriori algotirmasının geniş nesne kümelerinin ortaya çıkarılması ile ilgili ağaç diyagram sunulmuştur.



Şekil 2.4: Apriori Algoritması Ağaç Diyagram (Gündüz, 2015)

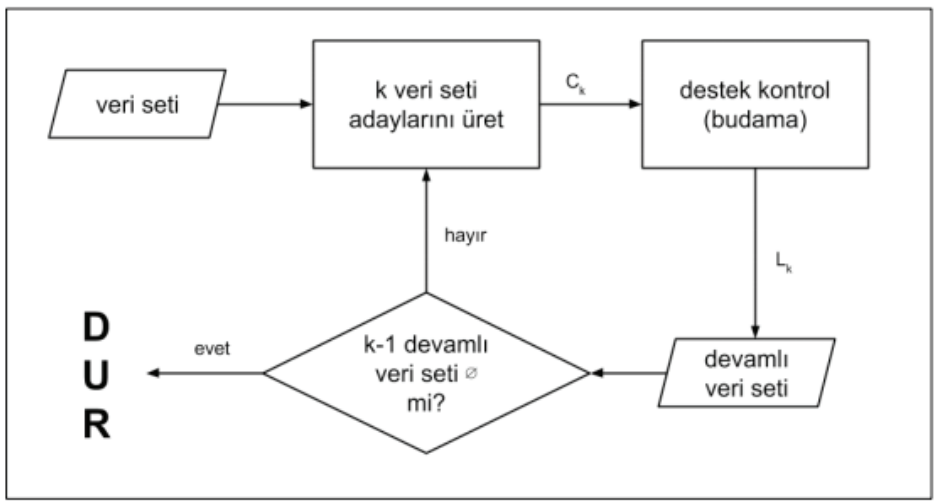
"Apriori, boolean ilişki kuralları için geçerli bir veri madenciliği algoritmasıdır" (Özçakır, 2006).

Apriori algoritmasının sözde kodu Şekil 2.5'de belirtilmiştir. Bu algoritmanın 1994 yılında Agrawal ve Srikant tarafından 20. Very Large Database Endowment konferansında sunulmuştur (Agrawal ve Srikant, 1993).

$$
\begin{aligned}
&\text{Apriori}(T, \epsilon)\\
&\quad L_1 \leftarrow \{\text{large } 1-\text{itemsets}\}\\
&\quad k \leftarrow 2\\
&\qquad \textbf{while } L_{k-1} \neq \ emptyset\\
&\qquad\quad C_k \leftarrow \{a \cup \{b\} \mid a \in L_{k-1} \wedge b \in \bigcup L_{k-1} \wedge b \notin a\}\\
&\qquad\quad \textbf{for } \text{transactions } t \in T\\
&\qquad\qquad C_t \leftarrow \{c \mid c \in C_k \wedge c \subseteq t\}\\
&\qquad\qquad \textbf{for } \text{candidates } c \in C_t\\
&\qquad\qquad\quad count[c] \leftarrow count[c] + 1\\
&\qquad\quad L_k \leftarrow \{c \mid c \in C_k \wedge \ count[c] \geq \epsilon\}\\
&\qquad\quad k \leftarrow k + 1\\
&\qquad \textbf{return } \bigcup_k L_k
\end{aligned}
$$

Şekil 2.5: Apriori Algoritmasının Sözde Kodu (Eker, 2016)

Şekil 2.5'de belirtilen apriori algoritmasının sözde kodunda anlatılmak istenen k-öğeli küme eğer minimum destek kriterini sağlıyor ise bu kümenin alt kümeleri de bu destek kriterlerini sağlamaktadır. Bir öğeler kümesindeki destek değeri alt kümesindeki destek değerinden büyük olmamaktadır. Sık gecen nesne kümesi altındaki altkümelerin tamamı boş olmaması durumunda altkümeleri sık geçmektedir. Bu özellik şu gözleme dayanmaktadır. Eğer bir nesne küme I, minimum destek eşik değeri olan minumum güven değerini sağlayamıyor ise, o zaman I sık geçen değildir denir (Eker, 2016). Apriori algoritmasının akış diyagramı Şekil 2.6'daki gibidir.



Şekil 2.6: Apirio Algoritması Akış Diyagramı (Eker, 2016)

Çok küçük bir veritabanından bile mümkün olan çok sayıda kural vardır, bu yüzden ilginç olanları seçmek için çeşitli ilgi ve öneme sahip tedbirler üzerinde kısıtlamalar kullanırız. Destek, güven, kaldırma ve mahkumiyet gibi yararlı önlemlerin bazılarıdır (Jain, 2017).

3. UYGULAMA

3.1. Apriori Algoritması ile Uygulama

Verilerin hazırlanması başlığında belirtilen işlemlerin gerçekleşmesinden sonra 2016 yılı yaz ve kış dönemlerine ait maskelenmiş uçuş biletleme verilerini Birleştirme kulları ve Apriori algoritması kullanılarak, minumum destek (min_support=0.1) değeri 0.1, lift (min_threshold=1) değeri 1 verilerek analiz edilmiştir. Belirtilen support değerini aşan 205 farklı grup ortaya çıkarılmıştır. Destek değeri oranları incelendiği zaman minimum %10.1 maksimum ise %47.6 analiz sonuçlarına ulaşılmıştır.

Uygulamada kullandığımız apriori algoritma kodu hazır yazılmış bir paket koddur. Kodun içerisin incelediğimiz zaman öncelikle parametreler tanımlanmıştır. Bazı parametrelerin default değerleri girilmiştir. Daha sonra veri türünün binary (0 ,1) olup olmadığı kotrol edilmiştir. Eğer veri türü binary değil ise, hata verilmesi sağlanmıştır. Bütün veri üzerine döngü kurularak her bir eşsiz değer için support hesaplanması

sağlanmıştır. Daha sonra elde ettiği eşsiz değerleri support ve itemsets kolanlarına sahip data frame üzerine yazdırılmıştır. Elde edilen supports değerleri birliktelik kuralı koduna koyularak güven ve lift değerlerinin hesaplanması sağlanmıştır.

Birliktelik kuralı kodunda ilk öncelikle foksiyonun içine parametlerin tanımlaması yapılmıştır. Birliktelik kuralında support ve itemsets kolonlarının olmasını zorunlu tutmaktadır. Destek, güven, lift ve diğer parametler hesaplanarak ilgili kolonlara aktarılması sağlanmıştır. Veri setinin analizi sonucu toplamda 824 adet yorumlanmak üzere kural ortaya çıkarılmıştır.

Uygulama sonucu elde edilen sonuçlarda öne çıkan kurallar, kış döneminde erkek yolcu tipinin iç hat uçuşlar için ekonomi sınıfını %96 güven, %12 destek değeri ile ekonomi sınıfı tercih ettiği Ayrıca 1,13 lift değeri 1'den büyük olması nedeniyle bağımsız olduğunu ortaya çıkarmıştır. Yine bir başka öne çıkan kuralda ise, yaz döneminde acentelerden alınan gidiş dönüş ekonomi sınıfı biletlerin %99 güven, %10 destek değeri ile uluslararası uçuşlarda alındığı ayrıca 1.84 lift değeri 1'den büyük olması nedeniyle bu olayların yüksek olasılıkla bağımsız olaylar olduğunu göstermiştir. Veri setinin analizi sonucu toplamda 824 adet yorumlanmak üzere kural ortaya çıkarılmıştır. Ortaya çıkarılan kuralların bir bölümü Şekil 3.1'de sunulmuştur.

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction
0	(SEASON_Summer)	(DOM-INT_International)	0.508603	0.542750	0.307692	0.604974	1.114646	0.031647	1.157519
1	(DOM-INT_International)	(SEASON_Summer)	0.542750	0.508603	0.307692	0.566912	1.114646	0.031647	1.134636
2	(SEASON_Summer)	(FLIGHT TYPE_RT)	0.508603	0.578761	0.319535	0.628261	1.085527	0.025176	1.133157
3	(FLIGHT TYPE_RT)	(SEASON_Summer)	0.578761	0.508603	0.319535	0.552102	1.085527	0.025176	1.097119
4	(SEASON_Summer)	(CABIN CLASS_Y)	0.508603	0.936522	0.476691	0.937255	1.000783	0.000373	1.011680
5	(CABIN CLASS_Y)	(SEASON_Summer)	0.936522	0.508603	0.476691	0.509001	1.000783	0.000373	1.000811
6	(CHANNEL_GDS)	(SEASON_Summer)	0.263491	0.508603	0.155608	0.590562	1.161146	0.021596	1.200176
7	(SEASON_Summer)	(CHANNEL_GDS)	0.508603	0.263491	0.155608	0.305951	1.161146	0.021596	1.061178
8	(SEASON_Summer)	(PASSENGER TITLE_MRS)	0.508603	0.170299	0.103332	0.203169	1.193015	0.016718	1.041251
9	(PASSENGER TITLE_MRS)	(SEASON_Summer)	0.170299	0.508603	0.103332	0.606771	1.193015	0.016718	1.249646
10	(PASSENGER TITLE_MS)	(SEASON_Summer)	0.283723	0.508603	0.163791	0.577291	1.135053	0.019488	1.162495
11	(SEASON_Summer)	(PASSENGER TITLE_MS)	0.508603	0.283723	0.163791	0.322041	1.135053	0.019488	1.056519
12	(ARRIVAL CITY_ISTANBUL)	(SEASON_Summer)	0.408384	0.508603	0.211697	0.518377	1.019218	0.003992	1.020295
13	(SEASON_Summer)	(ARRIVAL CITY_ISTANBUL)	0.508603	0.408384	0.211697	0.416232	1.019218	0.003992	1.013444
14	(SEASON_Summer)	(DEPARTURE CITY_ISTANBUL)	0.508603	0.422582	0.221205	0.434927	1.029214	0.006279	1.021847
15	(DEPARTURE CITY_ISTANBUL)	(SEASON_Summer)	0.422582	0.508603	0.221205	0.523461	1.029214	0.006279	1.031179
16	(DOM-INT_Domestic)	(SEASON_Winter)	0.457250	0.491397	0.256338	0.560609	1.140848	0.031647	1.157519
17	(SEASON_Winter)	(DOM-INT_Domestic)	0.491397	0.457250	0.256338	0.521652	1.140848	0.031647	1.134636
18	(FLIGHT TYPE_OW)	(SEASON_Winter)	0.419922	0.491397	0.231750	0.551888	1.123099	0.025401	1.134990
...	...	...	...	...	...	...	...	...	...
810	(FLIGHT TYPE_RT, DOM-INT_International)	(SEASON_Summer, CABIN CLASS_Y, DEPARTURE CITY_...	0.393852	0.205051	0.106894	0.271407	1.323605	0.026134	1.091074
811	(SEASON_Summer, FLIGHT TYPE_RT)	(DEPARTURE CITY_ISTANBUL, CABIN CLASS_Y, DOM-I...	0.319535	0.241736	0.106894	0.334531	1.383870	0.029651	1.139443
812	(CABIN CLASS_Y, FLIGHT TYPE_RT)	(DEPARTURE CITY_ISTANBUL, SEASON_Summer, DOM-I...	0.535524	0.150899	0.106894	0.199607	1.322784	0.026084	1.060855
813	(FLIGHT TYPE_RT, DEPARTURE CITY_ISTANBUL)	(SEASON_Summer, CABIN CLASS_Y, DOM-INT_Interna...	0.258551	0.280941	0.106894	0.413436	1.471611	0.034257	1.225883
814	(SEASON_Summer, DOM-INT_International)	(CABIN CLASS_Y, FLIGHT TYPE_RT, DEPARTURE CITY...	0.307692	0.236804	0.106894	0.347407	1.467069	0.034032	1.169484
815	(CABIN CLASS_Y, DOM-INT_International)	(SEASON_Summer, FLIGHT TYPE_RT, DEPARTURE CITY...	0.492590	0.145763	0.106894	0.217005	1.488755	0.035093	1.090987
816	(DEPARTURE CITY_ISTANBUL, DOM-INT_International)	(CABIN CLASS_Y, SEASON_Summer, FLIGHT TYPE_RT)	0.266995	0.296100	0.106894	0.400360	1.352109	0.027837	1.173871
817	(SEASON_Summer, CABIN CLASS_Y)	(DEPARTURE CITY_ISTANBUL, FLIGHT TYPE_RT, DOM-...	0.476691	0.194867	0.106894	0.224243	1.150748	0.014003	1.037867
818	(SEASON_Summer, DEPARTURE CITY_ISTANBUL)	(CABIN CLASS_Y, FLIGHT TYPE_RT, DOM-INT_Intern...	0.221205	0.356264	0.106894	0.483236	1.356400	0.028087	1.245707
819	(CABIN CLASS_Y, DEPARTURE CITY_ISTANBUL)	(SEASON_Summer, FLIGHT TYPE_RT, DOM-INT_Intern...	0.390245	0.237325	0.106894	0.273916	1.154182	0.014280	1.050395
820	(FLIGHT TYPE_RT)	(DEPARTURE CITY_ISTANBUL, SEASON_Summer, CABIN...	0.578761	0.137644	0.106894	0.184695	1.341836	0.027232	1.057710
821	(DOM-INT_International)	(CABIN CLASS_Y, SEASON_Summer, FLIGHT TYPE_RT,...	0.542750	0.134057	0.106894	0.196949	1.469143	0.034135	1.078316
822	(SEASON_Summer)	(DEPARTURE CITY_ISTANBUL, CABIN CLASS_Y, FLIGH...	0.508603	0.176058	0.106894	0.210173	1.193769	0.017351	1.043192
823	(DEPARTURE CITY_ISTANBUL)	(CABIN CLASS_Y, SEASON_Summer, FLIGHT TYPE_RT,...	0.422582	0.216280	0.106894	0.252955	1.169574	0.015498	1.049094

Şekil 3.1: Kurallar

4. SONUÇ VE ÖNERİLER

Yapılan bu çalışmada Türk Hava Yolları'nın 2016 yılı yaz ve kış dönemlerine ait verilerine Birliktelik kuralları ve Apriori algoritması uygulanmış ve çıkarılan ilişkili kurallar yorumlanarak güven değerlerinin yüksek olduğu kurallardan bir kısmı sunularak önerilerde bulunulmuştur.

Yaz döneminde acentelerden alınan gidiş dönüş ekonomi sınıfı biletlerin %99 güven değeri ile uluslararası uçuşlarda alındığı gözlemlenmiştir.

Bu kurala ilişkin, yaz döneminde yolcular yurt dışı seyehatlarini gidiş dönüş tarihi planlı olarak, ekonomik sınıf ile uygun fiyat gözeterek, mobil ve web satış kanallarını tercih etmeden doğrudan acente üzerinden satın almıştır. Bu kural sonucu ekonomi sınıfı daha çok tercih edildiğinden, bu kabin sınıfı özelinde uçuş deneyimini olumlu yönde arttıracak çalışmalar yapılabilir. Ayrıca biletleme işlemlerinin büyük bir bölümü acenteler üzerinden yapıldığından, acente personeli eğitimi, acentelerin genel durumlarında iyileştirme çalışmaları yapılabilinir. Kural sonucu bu yolcuların genellikle gidiş dönüş bilet tercih ettikleri görülmüştür.

Bu kurala dayanarak ,sadece gidiş bileti alan yolcular için dönüş uçuş bilet bilgileri de bu yolculara farklı kanallara aracılığıyla (acente personeli, e-mail vb.) sunulabilir.

Kış döneminde erkek yolcu tipinin iç hat uçuşlar için ekonomi sınıfını %96 güven değeri ile tercih ettiği gözlemlenmiştir. Bu kural fiyat belirleme ve kampanya stratejilerinde kullanılabilir. Yani kurala göre bu dönemde ekonomi sınıfına talep fazla olduğundan business sınıf, diğer yolcu tipleri ve dış hat uçuşlar ile ilgili kampanyalar yapılabilir.

Yapılan analiz sonuçları incelendiğinde tüm yolcu tipleri için satış platformlarında Mobil ve Web satış kanalları üzerinden yapılan biletlemelerin başka ürün grupları ile birlikte belirlenen destek değerini aşamadığı ve bununla ilgili kuralların oluşmadığı ortaya çıkmıştır. Bu tespit sonucu Türk Hava Yolları'nın mobil ve web satış platformlarının genel satış platformlarının çok altında kalması bu platformlara yapılacak geliştirmeler ya da özel kampanyalar ile müşterinin ilgisini çekebilecek stratejiler belirlenebilir.

Yapılan kural analizleri ve yorumları sonucu bu metodun stratejik kararlar, kampanyalar, geliştirmeler ve personel eğitimi gibi konularda etkili olabileceği gösterilmiştir. Sonraki yapılacak çalışmalarda, dağıtık sistemlerde birliktelik kuralı algoritmaları da uygulanabilir.

5. KAYNAKLAR

Agrawal, R., Srikant, R. (1993). Fast Algorithms For Mining Association Rules, Conference on Very Large Databases, Santiago, Chile. 487-499.

Akpınar, H. (2000). Veri Tabanlarında Bilgi Keşfi ve Veri Madenciliği, İstanbul Üniversitesi, İşletme Fakültesi Dergisi, C.XXIX, No.1, s.1-22.

Al-Rubaiee, B. (2018). Data Mining and an Application in The Open Education System of Anadolu University, Master Thesis, Graduate School of Sciences.

Altun, M. (2017). Veri Madenciliği ve Uygulama Alanları, Doktora Semineri Raporu, Akdeniz Üniversitesi, Eğitim Bilimleri.

Berry, M. J. A. ve Linoff, G. S. (2004). Data Mining Techniques for Marketing, Sales and Customer Relationship Management, Wiley Publishing, Inc., Indianapolis.

Döşlü A. (2008). Veri Madenciliğinde Market Sepet Analizi ve Birliktelik Kurallarının Belirlenmesi, Yüksek Lisans Tezi, Yıldız Teknik Üniversitesi, Fen Bilimleri Enstitüsü.

Eker, M. E. (2016). Veri Madenciliğinde Apriori Algoritmasının Sınav Verileri Üzerinde Uygulanması, Yüksek Lisans Tezi, Ondokuz Mayıs Üniversitesi, Fen Bilimleri Enstitüsü.

Farboudi, S. (2009). Tıp Bilişiminde İstatistiksel Veri Madenciliği, Yüksek Lisans Tezi, Hacette Üniversitesi, Fen Bilimleri Enstitüsü.

Fayyad, U., Piatetsky-Shapiro, G., Smyth, P. (1996). From Data Mining To Knowledge Discovery In Databases, AI Magazine, sayı 17, syf. 37-54.

Gülce, A. C. (2010). Veri Medenciliğinde Apriori Algoritması ve Apriori Algoritmasının Farklı Veri Kümelerinde Uygulanması, Yüksek Lisans Tezi, Trakya Üniversitesi, Fen Bilimleri Enstitüsü.

Jacobs, P. (1999). Data Mining: What General Managers Need to Know, Harvard Management Update 4, 8.

Jain, R. (2017). Applıcatıon Of Data Mınıng Technıques For Predıctıng Students Academic Performance: A Study, Internatıonal Journal of Innovation in Engineering Research & Management, Vol. 04, No.3, June 2017.

Kılınç, Y. (2009). Mining Association Rules For Quality Related Data In An Electronics Company,Master Thesis, Middle East Technical University, Industrial Engineering.

Lai, C., Lu, Jin. (2018). Evaluating the Efficiency of Currency Portfolios Constructed By the Mining Association Rules, Asia Pacific Management Review, Vol. 23, Issue 3, pp.161-234

Öğüt, M. (2005). Örneklere Dayalı Bir Sınıflandırma Algoritma Tasarımı ve Uygulaması, Yüksek Lisans Tezi, Selçuk Üniversitesi, Fen Bilimleri Enstitüsü.

Özçakır F. C. (2006), Müşteri İşlemlerindeki Birlikteliklerin Belirlenmeside Veri Madenciliği Uygulaması, Yüksek Lisans Tezi, Marmara Üniversitesi, Fen Bilimleri Enstitüsü.

Özkan, Y. (2008). Veri Madenciliği Yöntemleri, Papatya Yayıncılık Eğitim.

Sadıqmal F. (2015). Implementation of Some Medical Data Using Apriori Algorithm, Master Thesis, Sakarya University, Institute Of Scıence And Technology.

Sönmez, F. (2018). Anomaly Detection Using Data Mining Methods in IT Systems: A Decision Support Application, Sakarya University Journal Of Science, 22(4): 1190-1123

Ulaş, M. A. (1999). Market Basket Analysis For Data Mining, Master Thesis, Bogaziçi University, Computer Engineering.

Xu, C., Bao, J., Wang, C., Liu, P. (2018). Association Rule Analysis of Factors Contributing to Extraordinarily Severe Traffic Crashes in China, Journal of Safety Research, Volume 67, Pages 65-75.

Yıldız, B. (2010). Impacts Of Frequent Itemset Hiding Algorithms On Privacy Preserving Data Mining, Master Thesis, İzmir İnstitute of Technology, Computer Engineering.

Zaimoğlu, E. A. (2018). Veri Madenciliği Teknikleri Kullanılarak Sosyal Ağlar Aracılığı İle Bilgisayar ve Bilişim Mühendisliği Mezun Öğrenci Profillerinin Belirlenmesi, Yüksek Lisans Tezi, Sakarya Üniversitesi, Fen Bilimleri Enstitüsü.