

T.C.
İSTANBUL AYDIN ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ



VERİ MADENCİLİĞİ YÖNTEMLERİ İLE UÇUŞ BİLETLEME ANALİZİ

YÜKSEK LİSANS TEZİ

Muhammed Metin ULUYARDIMCI

Bilgisayar Mühendisliği Ana Bilim Dalı

Bilgisayar Mühendisliği Programı

ARALIK 2018

T.C.
İSTANBUL AYDIN ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ



VERİ MADENCİLİĞİ YÖNTEMLERİ İLE UÇUŞ BİLETLEME ANALİZİ

YÜKSEK LİSANS TEZİ

Muhammed Metin ULUYARDIMCI

(Y1513.010039)

Bilgisayar Mühendisliği Ana Bilim Dalı

Bilgisayar Mühendisliği Programı

Tez Danışmanı: Doç. Dr. Metin ZONTUL

ARALIK 2018



T.C.
İSTANBUL AYDIN ÜNİVERSİTESİ
FEN BİLİMLER ENSTİTÜSÜ MÜDÜRLÜĞÜ

Yüksek Lisans Tez Onay Belgesi

Enstitümüz Bilgisayar Mühendisliği Ana Bilim Dalı Bilgisayar Mühendisliği Tezli Yüksek Lisans Programı Y1513.010039 numaralı öğrencisi **M. Metin ULUYARDIMCI**'nin “**VERİ MADENCİLİĞİ YÖNTEMLERİ İLE UÇUŞ BİLETLEME ANALİZİ**” adlı tez çalışması Enstitümüz Yönetim Kurulunun 29.11.2018 tarih ve 2018/24 sayılı kararıyla oluşturulan jüri tarafından *o.g. b.ö.ö.ö.* ile Tezli Yüksek Lisans tezi olarak *kabul* edilmiştir.

Öğretim Üyesi Adı Soyadı

İmzası

Tez Savunma Tarihi : 26/12/2018

1) Tez Danışmanı: Doç. Dr. Metin ZONTUL

2) Jüri Üyesi : Prof. Dr. Ali GÜNEŞ

3) Jüri Üyesi : Dr. Öğr. Üyesi Ferdi SÖNMEZ

Metin Zontul
.....
Ali Güneş
.....
Ferdi Sönmez
.....

Not: Öğrencinin Tez savunmasında **Başarılı** olması halinde bu form **imzalanacaktır**. Aksi halde geçersizdir.

YEMİN METNİ

Yüksek Lisans tezi olarak sunduđum “**VERİ MADENCİLİĐİ YÖNTEMLERİ İLE UÇUŞ BİLETTLEME ANALİZİ**” adlı alıřmanın, tezin proje safhasından sonuçlanmasına kadarki bütün süreçlerde bilimsel ahlak ve geleneklere aykırı düşecek bir yardıma başvurulmaksızın yazıldığını ve yararlandığım eserlerin Bibliyografya’da gösterilenlerden oluştuđunu, bunlara atıf yapılarak yararlanılmış olduğunu belirtir ve onurumla beyan ederim. (26.12.2018)

Muhammed Metin ULUYARDIMCI

*Bu alıřmayı, hazırlık srecinde
beni destekleyen Dogan Can DEMİRBILEK ve
Eřime, Annem ve Babama ithaf ediyorum.*

ÖNSÖZ

Yüksek lisans tezim boyunca benden yardımlarını esirgemeyen, karşılaştığım sorunlarda bilgi ve deneyimlerini benimle paylaşan değerli hocam ve tez danışmanım Sayın Doç. Dr. Metin ZONTUL'a, çalışmamda destekleri olan Sayın Dr.Öğr. Üyesi Ferdi SÖNMEZ, Sayın Veysel Çağın KARADENİZ ve Türk Hava Yolları Ar-Ge Merkezi'ne teşekkürü bir borç bilirim.

Son olarak, hayatımın her döneminde olduğu gibi tez çalışmalarım boyunca da destek olan aileme sonsuz teşekkür eder, saygılarımı sunarım.

Aralık 2018

Muhammed Metin ULUYARDIMCI

İÇİNDEKİLER

| | <u>Sayfa</u> |
|--|--------------|
| ÖNSÖZ..... | ix |
| İÇİNDEKİLER | xi |
| ÇİZELGE LİSTESİ..... | xiii |
| ŞEKİL LİSTESİ..... | xv |
| ÖZET..... | xvii |
| ABSTRACT..... | xix |
| 1. GİRİŞ | 1 |
| 1.1 Literatür Taraması..... | 2 |
| 2. VERİ MADENCİLİĞİ | 9 |
| 2.1 Veri Madenciliği ve Bilgi Keşfi..... | 10 |
| 2.2 Veri Madenciliğinin Gelişimi | 11 |
| 2.3 Veri Madenciliği Yöntemleri | 12 |
| 2.3.1 Öngörü yöntemi | 12 |
| 2.3.2 Tanımlayıcı model | 12 |
| 2.4 Veri Madenciliği Uygulama Alanları | 13 |
| 2.5 Birliktelik Kuralları | 14 |
| 2.5.1 Güven destek ve diğer kavramları | 15 |
| 2.6 Apriori Algoritması..... | 17 |
| 3. MATERYAL VE YÖNTEM..... | 23 |
| 3.1. Kullanılan Teknolojiler..... | 23 |
| 3.1.1. Anaconda | 23 |
| 3.1.2 Python | 23 |
| 3.1.3. One-Hot encode | 24 |
| 3.1.4. Bar chart (Çubuk Grafiği)..... | 25 |
| 3.1.5. Heat map (Isı Haritası)..... | 25 |
| 3.2. Verilerin Hazırlanması | 25 |
| 3.2.1. Veri temizleme..... | 27 |
| 3.2.2. Veri entegrasyonu | 27 |
| 3.2.3. Veri seçimi | 28 |
| 3.2.4. Veri dönüşümü..... | 28 |

| | |
|---|-----------|
| 3.3. Verilerle İlgili Tanım ve İstatistik Bilgiler | 29 |
| 3.4. Apriori Algoritması ile Uygulama..... | 35 |
| 4. SONUÇ VE ÖNERİLER..... | 41 |
| KAYNAKLAR..... | 43 |
| ÖZGEÇMİŞ..... | 47 |

ÇİZELGE LİSTESİ

| | <u>Sayfa</u> |
|---|---------------------|
| Çizelge 2. 1: Alışveriş Sepeti..... | 20 |
| Çizelge 2. 2: Birinci Destek Değer Taraması | 21 |
| Çizelge 2. 3: İkinci Destek Değer Taraması..... | 21 |
| Çizelge 2. 4: Üçüncü Destek Değer Taraması..... | 21 |
| Çizelge 2. 5: Maximum Kümeler ve Destek, Güven Yüzdeleri | 22 |

ŞEKİL LİSTESİ

Sayfa

| | |
|--|----|
| Şekil 2.1: Veri Madenciliği Süreci (Özkan, 2008)..... | 11 |
| Şekil 2.2: Veri Madenciliği Tanımlayıcı ve Öngörü Yöntemleri (Zaimoğlu, 2018)..... | 14 |
| Şekil 2.3: Ürünler Satış..... | 17 |
| Şekil 2.4: Apriori Algoritması Ağaç Diyagram (Gündüz, 2015)..... | 19 |
| Şekil 2.5: Apriori Algoritmasının Sözde Kodu (Eker, 2016)..... | 20 |
| Şekil 2.6: Apriori Algoritması Akış Diyagramı (Eker, 2016)..... | 21 |
| Şekil 3.1: Maskelenmiş Uçuş Biletleme Verileri - 1..... | 28 |
| Şekil 3.2: Maskelenmiş Uçuş Biletleme Verileri - 2..... | 29 |
| Şekil 3.3: Veri Birleştirme Kodu ve Veri Seti..... | 30 |
| Şekil 3.4: Verilerin Dönüştürülmesi..... | 31 |
| Şekil 3.5: Şehirlere Göre Uçuş Sayısı..... | 32 |
| Şekil 3.6: Satış Kanalları..... | 33 |
| Şekil 3.7: Satış Kanalları Üzerinden Yapılan Biletleme Sayısı..... | 34 |
| Şekil 3.8: Yolcu Ünvanlarının Tanımları..... | 34 |
| Şekil 3.9: Yolcu Ünvanları Göre Yurt İçi ve Yurt Dışı Biletleme Sayıları..... | 35 |
| Şekil 3.10: Kabin Tipleri | 35 |
| Şekil 3.11: Uçuş Tipi Analizi Sonucu..... | 36 |
| Şekil 3.12: Veri Seti Analiz Grafiği..... | 36 |
| Şekil 3.13: Uygulamadaki Apriori Kodu..... | 38 |
| Şekil 3.14: Support Değerini Aşan Gruplar..... | 39 |
| Şekil 3.15: Oluşan Kurallar | 40 |
| Şekil 3.16: İsim Haritası..... | 41 |
| Şekil 3.17: Çubuk Grafiği..... | 42 |

VERİ MADENCİLİĞİ YÖNTEMLERİ İLE UÇUŞ BİLETLEME ANALİZİ

ÖZET

Bir çok sektörde kullanılan veri madenciliği hava yolu şirketleri açısından 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. 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 yöneltmiştir. 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 haline almıştır. Artık alışveriş, eğlence, ulaşım gibi bir çok ihtiyaç, 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 nesnelere arası ilişkileri ortaya çıkarmak için kullanılır (Eker, 2016). Bu veriler müşterilerin yapmış olduğu uçuşlarda tercih ettikleri bir çok faktörün göz önüne alınarak gelecekte 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 ile, bir çok kurum ve kuruluş kendi amaçları doğrultusunda 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).

Veri madenciliği yöntemlerinin kullanıldığı bu çalışmada, ilişkisel veri tabanları üzerinde çoklu ilişkisel yapıdaki ortak kuralları bulmayı sağlayan bir uygulama geliştirilmiştir. Uygulama altyapısı olarak ilişkisel veri tabanlarındaki desenleri tanımlayabilen, bu desenleri eklerle geliştirebilen ve bu desenlerin çeşitli ölçmeleri için gerekli sayımları veri tabanından temel yetilerle alan bir yapı kullanılmıştır. Bu altyapı, veri tabanının tanımında yer alan bilgileri kullanarak arama alanının daraltılmasını sağlamıştır. Bu çalışma, Apriori algoritmasını arama alanını daha da küçültmek için kullanarak ve altyapı tarafından desteklenmeyen özyinelemeli desenlerin bulunmasını sağlayarak altyapıya yenilikler getirmiştir. Apriori algoritması her tablo üzerinde sık karşılaşılan desenleri bulmak için kullanılmış ve bu algoritmanın gerekli destek değerini bulma yöntemi değiştirilmiştir. Veri tabanındaki özyinelemeli ilişkileri belirlemek için bir yöntem sunulmuş ve uygulama bu durumlar için tablo kısaltmalarının kullanıldığı bir çözüm sağlamıştır. Veri tabanı alanlarında saklanan sürekli değerleri bölümleyebilmek için eşit derinlik yöntemi kullanılmıştır. Uygulama bir veri madenciliği yarışması olan KDD Cup 2001'den alınan örnek genlerde yer tahmini problemi ile test edilmiş ve ortaya çıkan sonuçlar yarışmayı kazanan yaklaşımın sonuçlarıyla karşılaştırılmıştır (Toprak, 2004).

Yapılan bir makele çalışmasında, bilişim, iletişim ve alt yapı teknolojisinin yüksek entegrasyonu sebebiyle kontrol sistemlerinden beklenmedik durumlar ile karşılaşıldığını belirtilmiş ve bir yöntem geliştirilmiştir. Bu çalışmada kontrol sistemlerinin dinamik davranışları veri madenciliği yöntemlerinden biri olan birliktelik kuralı kullanılarak analiz edilmiş ve kontrol sistemlerinin normal davranış özelliklerini özetleyen bir veri madenciliği yöntemi önerilmiştir. Kontrol sistemlerinde sık kullanılan davranışlar ile normal kullanılan davranışlar karşılaştırılmasından yola çıkılarak bir güvenlik parametresi eklenerek kontrol sisteminin anormal beklenmedik davranışlarının zamanında tespit

edildiğini gösterilmiştir (Jie ve arkadaşları, 2018).

Eyüp Sıramkaya 2005 yılında yapmış olduğu çalışmada, geliştirilen bir uygulama ile sanal ortamlardaki basın ve yayın kaynaklarının metinsel ve görsel olarak verileri analiz edilerek istihbarat tarafından önemli olan olaylar ve kişiler ile ilgili kaynaklara yoğunlaşarak verilerin hızlı ve etkin bir şekilde ulaşılması ve önemli bilgiler elde edilmesi amaçlanmıştır. İnternet üzerinden erişilen haber kaynakları yazılımlar ile veri tabanına işlenerek depolanması sağlanmıştır. Depolanan verilere istenildiği zaman ulaşılması için bir arayüz geliştirmiştir. Bu çalışmanın analizi aşamasında veri madenciliğinde sıkça kullanılan birliktelik kuralları ve apriori algoritması kullanılmıştır. Kişi-kişi ilişkilerinin ortaya çıkarılması için bulanık mantık çalışması uygulanmıştır. Yapılan çalışmanın sonuçlarına göre "büyük-küçük harf" duyarlı bir klasik karşılaştırma algoritması ile "bulanık mantık" algoritması arasında çok büyük bir üstünlük olduğu tespit edilmiştir. Bununla birlikte lineer olarak benzerlik için kullanılan algoritmalarda da üstünlük sağladığı gösterilmiştir (Sarımkaya, 2005).

Seda Dağlar Toprak 2005 yılında hazırladığı yüksek lisans tez çalışmasında, yeni bir melez çok ilişkili veri madenciliği tekniği gerçekleştirilmiştir. Bu çalışmada kavram öğrenme, kavram ile kavramı gerçekleştirme önşartları arasındaki eşleştirme olarak belirtilmiş ve ilişkisel kural madenciliği alanında buluşsal yöntem olarak kullanılan Apriori algoritması örüntü uzayını azaltma amacı ile kullanılmıştır. Önerilen sistem, kavram örneklerinden ters çözünürlük operatörü kullanılarak genel kavram tanımlarını oluşturan ve bu genel örüntüleri Apriori algoritmasını temel alan bir operatör yardımı ile özelleştirerek güçlü kavram tanımlamaları elde eden melez bir öğrenme sistemi olarak belirtilmiştir. Sistemin iki farklı sürümü, üç popüler veri madenciliği problemi için test edilmiş ve sonuçlar önerilen sistemin, en gelişkin ilişkisel veri madenciliği sistemleri ile karşılaştırılabilir durumda olduğunu göstermiştir (Toprak, 2005).

Ayşe Onat 2008 yılında hazırladığı yüksek lisans çalışmasında da web madenciliği sayesinde müşterilerin ilgi alanları, ürün pazarlama stratejileri oluşturma, reklam alma, insanların birbirleriyle olan ilişkilerini gözden geçirme, insanların hangi sayfalarda daha çok gezdiğini belirleyerek o sayfalarda daha farklı sunumlar oluşturabilme gibi hususlarda kurumlara yardımcı olunması düşünülerek, sanal ortamlar üzerinden kişilerin yaşadığı

yer, yaş, cinsiyet, eğitim durumu gibi bir takım özelliklerini web madenciliği yöntemlerinden biri olan Apriori algoritması uygulayarak kişilerin birbirleriyle olan benzerlikleri ve uyumlarının keşfedilmesini amaçlamıştır (Onat, 2008).

Ufuk Ekim 2011 yılında hazırladığı yüksek lisans tez çalışmasında, Selçuk üniversitesi öğrenci işlerinde kullanılan otomasyon uygulamasından alınan veriler ile öğrencilerin gelecekteki durumları hakkında tahmin yapabilinmesi amaçlanmıştır. Apriori algoritması ve karar ağacı algoritması kullanılarak Selçuk Üniversitesi'ni yeni kazanmış bir öğrencinin üniversitedeki gelişimine etki eden başarı faktörleri araştırılmış ve yapılan çalışma sonucu ailelerin eğitim seviyesi ve gelir düzeyinin çocuklarının (öğrencinin) gelişimi ve başarısı üzerinde en etkili durumlar olduğu görülmüştür. (Ekim, 2011)

Yapılan bir başka çalışmada, veritabanlarında bilgi keşfi süreçleri, veri madenciliği, veri madenciliğinde kullanılan birliktelik-ilişki kuralı ve Apriori algoritması hakkında bilgiler verilmiştir. Uygulama bölümünde, gerçek veriler kullanarak Birliktelik Kuralları yöntemi ile Pazar Sepeti Çözümlemesi uygulaması yapılmış ve elde edilen sonuçlar tartışılmıştır (Şen, 2008).

Zahraa Mohammed Malık Malık 2018 yılında hazırladığı yüksek lisans tez çalışmasında, satın alınan miktarlara dayalı veri kümesinden otomatik olarak bilgi alınması ile veri kümesindeki ürünler arasındaki ilginç negatif birliktelik kurallarını bulma kabiliyetine sahip yeni bir teknik önerilmektedir. Birliktelik kuralı madenciliği, gözetimsiz veri madencilik tekniği olduğundan, sağlanan veri kümesi etiketsiz verilerden oluşmaktadır. Bu çalışmada, K-Ortalama(k means) ve Gürültülü Uygulamaların Yoğunluk Tabanlı Uzaysal Kümelmesi (DBSCAN) olmak üzere iki yöntem test edilmiştir. Bu tekniklerin sonuçları, herhangi bir alan bilgisi olmaksızın bulunan negatif birliktelik kuralı sonuçları ile karşılaştırılmıştır. DBSCAN kümeleme yönteminin kullanımı, gerçek yaşam işlem veri tabanında test edildiğinde %0.21 destek ve %91.84 güven ortalama değerleri ile 4,086 şeklinde daha iyi negatif birliktelik kuralı sonucu göstermiştir. K-Ortalama kümeleme yönteminin kullanımı ile çıkarılan alan bilgisine dayalı negatif birliktelik kuralları madenciliği sonucu, %0.19 destek ve %85.84 güven ortalama değerine sahip 1,780 iken, alan bilgisiz negatif birliktelik kuralı sonucu %0.12 destek ve %99.37 güven ortalama değerli 9,066 sonucunu vermiştir (Malık Malık, 2018).

Mine Durdu 2012 yılında hazırladığı yüksek lisans tez çalışmasında, perakende alanında yer alan bir firmanın müşteri bilgilerinin tutulduğu veri tabanı analiz edilerek elde edilen bilgi veya bilgilerden müşteri sadakatinin gelişmesi amaçlanmıştır. Veri tabanının analizi sürecinde veri madenciliği yöntemlerinden birliktelik kuralları kullanılmıştır. Veri madenciliğinde sık görünen nesnelere keşfedilmesinde kullanılan birliktelik kurallarından biri olan apriori algoritması uygulanmıştır. Sık geçen nesnelere ortaya çıkarılması için apriori algoritması birleştirme, eleme ve destek ölçütü kullanılarak birden fazla tarama yapılarak ilişkili nesnelere ortaya çıkarılmıştır. Analiz sonrası elde edilen sonuçlar oranlarla yorumlanmıştır (Durdu, 2012).

Seniha Gündüz 2017 yılında hazırladığı çalışmasında, Market Sepet Analizinde yaygın olarak kullanılan Apriori algoritması ile müşterilerin alışveriş bilgilerine göre birlikteliklerin ortaya çıkarılması amaçlanmaktadır. Verilerin analizinde SPSS Clementine paket programı kullanılmıştır. Elde edilen sonuçlara göre en fazla birlikte satın alınan ürün kategorileri belirlenmiş ve bu bilgiler ışığında market düzenlemelerinde, raf dizilimlerinde ve market yöneticilerinin promosyonlu ürünleri belirlemelerinde alternatif fikir ya da düşünceleri de dikkate alabilecekleri ortaya konmuştur (Gündüz, 2017).

Veri madenciliği yöntemlerinin kullanılarak 2017 yılında yapılan bir çalışmada pazarlama alanındaki bir firmanın müşterilerinin memnuniyetini sağlamak, satış oranlarını yükseltmek ve müşterinin firmaya olan bağlılığını artırmak amaçlanmıştır.

Yapılan bir çalışmada yüksek miktardaki satış verileri kullanılarak müşteri kitlelerinin hangi ürünleri satın aldıklarını ürünler arasındaki ilişkiyi birliktelik kuralları kullanılarak tespit edilmiştir. Elde edilen birliktelik kuralları sonucu müşteri kitlelerinin ürünler ile olan ilgilerine göre birden fazla kampanya oluşturularak müşterinin kendini özel hissetmesi amaçlanmıştır. Bununla birlikte müşteri için hazırlanan kampanyalardaki ürünlerin satış olasılıkları yapay sinir ağları yöntemi kullanılarak kampanyanın tutarlılığı hakkında önemli sonuçlara ulaşılmıştır. Yapay sinir ağları yöntemi ise MATLAB üzerinden gerçekleştirilmiştir. Uygulamada verilerin tutulması ve analize uygun bir hale getirilmesi için MSSQL Server sistemi kullanılmıştır. Birliktelik kuralları ile birlikte apriori algoritması kullanılarak veri analiz edilerek yorumlanmıştır.

Ortaya ıkartılan kuralları daha kolay ve anlaşılır olması iin gsterge panelleri tasarlanmıřtır (ztemiz, 2017).

Adnan Karaibrahimođlu ve Ařır Gen tarafından 2014 yılında yapılan alıřmada, Meram Tıp Fakltesi Onkoloji Hastanesine ait retrospektif alıřma sonucu ulařılan meme kanseri verilerinde Apriori algoritması kullanarak birliktelik rntlerini ortaya ıkarmıřlardır (Karaibrahimođlu ve Gen, 2014).

Feriřtah Dalkılı ve mer Aydın tarafından 2017 yılında yapılan bir alıřmada đrencilerin devamsızlık durumlarına gre bařarı durumlarının Apriori algoritması kullanılarak birliktelik kurallarının ortaya ıkarılmıřtır. Analiz sonularına gre, blm tr, cinsiyet đrenim tr, kayıtlı olunan yılın genel bařarı durumu ile birlikte devamsızlık yapmalarının zerinde etkileri olduđu tespit edilmiřtir (Dalkılı ve Aydın, 2017).

Bu tez alıřmasında iki milyon uuř biletleme verisi zerinde veri madenciliđi yntemlerinden biri olan birliktelik kuralları ve apriori algoritması Trkiye’de ilk defa uygulanmıřtır. alıřmanın ikinci blmnde kullanılan kuralların ve veri madenciliđin tanımlarına yer verilmiřtir. n blmde uygulama detaylı bir řekilde ele alınarak adım adım uygulanan iřlemler ve analiz sonuları ekran grntleri ile birlikte paylařılmıřtır. Sonu ve neriler kısmında analiz sonucu elde edilen kurallar yorumlanarak nerilerde bulunulmuřtur.

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 keşfederek 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 amaçlı 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).

“Veri madenciliği, büyük miktarlardaki veriyi bilgiye dönüştürmektir” (Han ve ark., 2011).

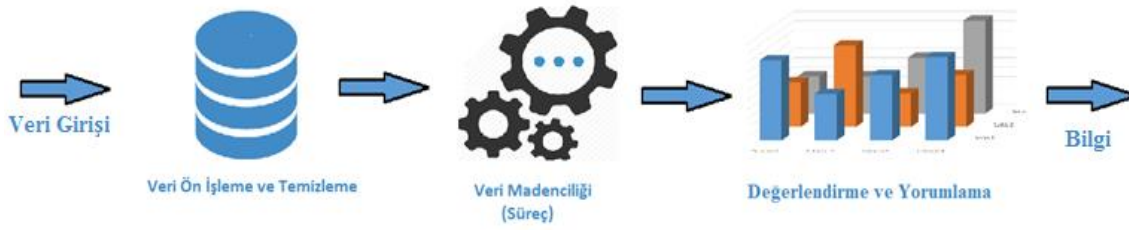
“Veri madenciliği, veri içerisindeki kullanışlı örüntülerin bulunması sürecidir” (Roiger ve Geatz, 2003).

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 uygulanarak yararlı bilgiler ayrıştırılır ve bu süreç veri madenciliği 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 amacıyla 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 sağlamıştır (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

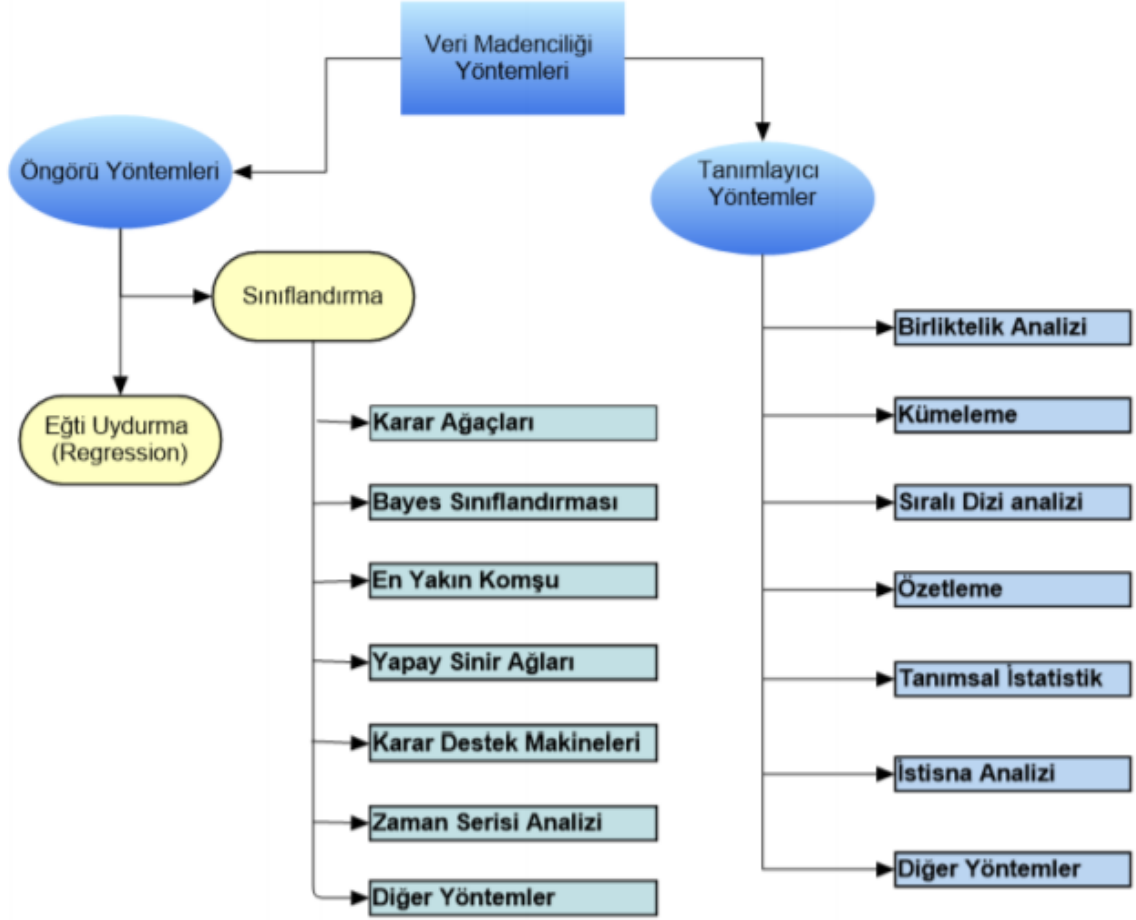
2.3.1 Öngörü yöntemi

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 kullanarak 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 çözümlene 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 (Zaimođlu, 2018)

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

2.4 Veri Madenciliđi Uygulama Alanları

Veri madenciliđi, 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 geçtikçe veri madenciliđi 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 aldığı anda %75 ihtimalle cips de 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 erişilebilmektedir. Amaç yararlı olan kuralları bulabilmektir. Yararlılığı ölçmenin yolu güven ve destek değerlerinden geçmektedir. Bu değerler niteliğe göre keşfedilen kuralların kullanılabilirliğini ve doğruluğunu ifade eder (Gülce, 2010).

Güven ve destek değerlerinin örnek bir formülü aşağıdaki gibidir:

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

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

$$\text{Güven} = \frac{\text{X ve Y'nin bulunduğu satır sayısı}}{\text{X'sin bulunduğu satır sayısı}} \quad (2.1)$$

Güven oranı %55 olan durumda; X ürünü almak isteyen müşterilerin %55'ini Y ürünüde 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 ürünüde alınması olayına, desteği ise aşağıdaki gibi hesaplanabilir (2.2):

$$\text{Destek} = \frac{\text{X ve Y'nin bulunduğu satır sayısı}}{\text{Toplam satır sayısı}} \quad (2.2)$$

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 ürünüde alınması arasındaki ilişki, aşağıdaki örnekle açıklanabilir.

$$\text{Güven} = \frac{(C, D \rightarrow A) \quad 3}{(C, D) \quad 5} = 0.6 \quad (2.3)$$

$$\text{Destek} = \frac{(C, D \rightarrow A) \quad 3}{\text{Toplam hareket} \quad 6} = 0.5 \quad (2.4)$$

Yapılan güven ve destek değerlendirmeleri sonucu (C, D) ürünlerinin yapıldığı satışlarda A ürünüde 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).

$$\text{Lift}(A \rightarrow C) = \frac{\text{Confidence}(A \rightarrow C)}{\text{Support}(C)}, \text{ aralık: } [0, \infty] \quad (2.5)$$

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).

$$\text{Levarage}(A \rightarrow C) = \text{support}(A \rightarrow C) - \text{support}(A) * \text{support}(C), \text{ aralık: } [-1,1] \quad (2.6)$$

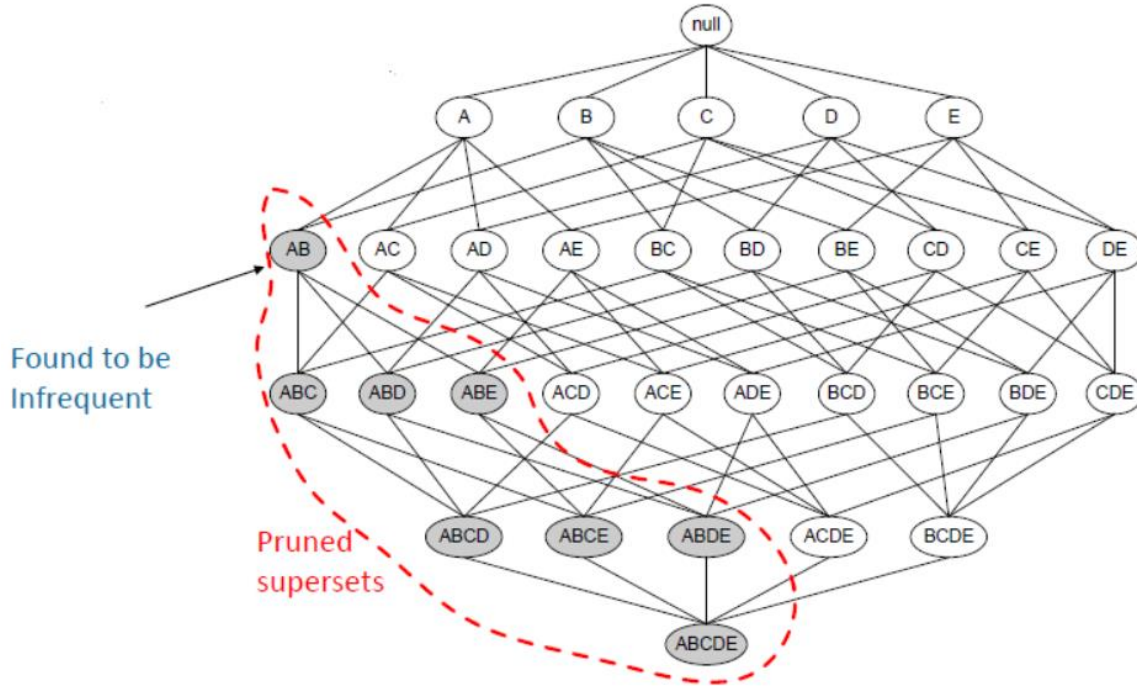
Conviction, A ve C'nin birlikte ya da A ürününün C ürünü olmaksızın görülme olasılıkları hesaplanması için kullanılır. Formülü aşağıdaki gibidir (Lia ve Lu, 2018).

$$\text{Conviction}(A \rightarrow C) = 1 - \frac{\text{Support}(C)}{1 - \text{Confidence}(A \rightarrow C)}, \text{ aralık: } [0, \infty] \quad (2.7)$$

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 ile 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 çıkarılarak 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 algoritmasının geniş nesne kümelerinin ortaya çıkarılması ile ilgili ağaç diyagramı sunulmuştur.



Şekil 2.4: Apriori Algoritması Ağaç Diyagramı (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).

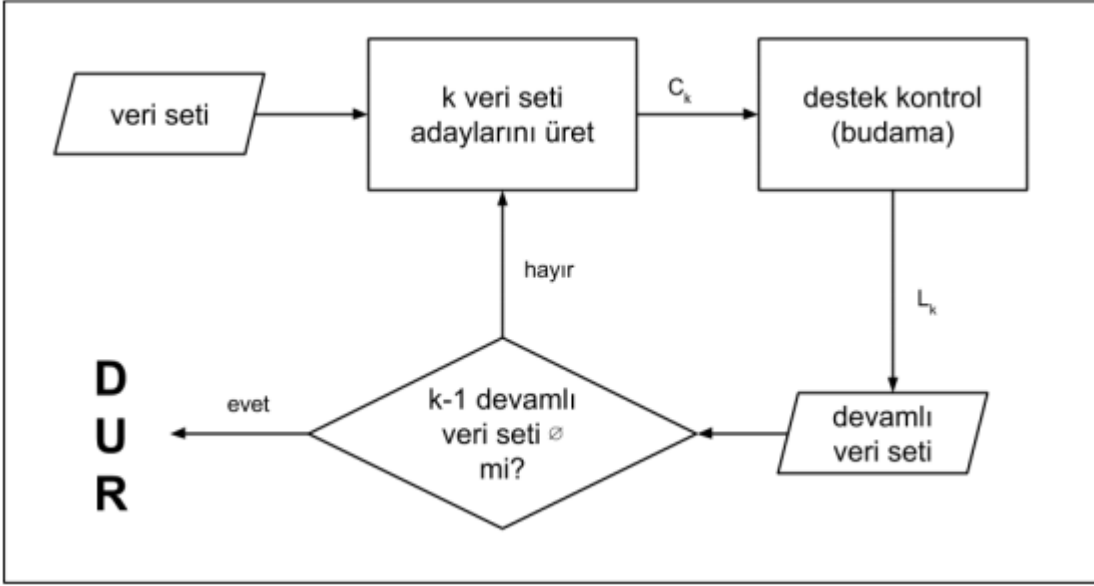

```

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

```

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 minimum güven değerini sağlayamıyor ise, o zaman I sık geçen değildir denir (Eker, 2016).



Şekil 2.6: Apriori 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 kullanılır. Destek, güven, kaldırma ve mahkumiyet gibi yararlı önlemlerin bazılarıdır (Jain, 2017).

Örnek

Çizelge 2.1'deki alışveriş listesindeki verileri kullanarak minimum güven %58 ve destek %27 olacak şekilde apriori algoritmasının kullanılması.

Çizelge 2. 1: Alışveriş Sepeti

| TID | ÜRÜNLER |
|-----|--------------------------------|
| 1 | Ekmek, Peynir, Salata, Domates |
| 2 | Peynir, Domates, Ekmek |
| 3 | Yumurta, Ekmek |
| 4 | Salata, Yumurta, Yağ |
| 5 | Peynir, Domates |
| 6 | Peynir, Domates, Yumurta |

Veri setinde içerisindeki farklı veri tipleri ortaya çıkarılarak veri tabanındaki miktarlarının ve veri tabanındaki toplam kayıt sayıları alınarak destek değerleri bulunur.

Çizelge 2. 2: Birinci Destek Değer Taraması

| Ürünler | Miktar | Destek Oranı (%) |
|---------|--------|------------------|
| Peynir | 4 | 67 |
| Ekmek | 3 | 50 |
| Salata | 2 | 33 |
| Domates | 4 | 67 |
| Yumurta | 3 | 50 |
| Yağ | 1 | 17 |

Çizelge 2.2'deki birinci destek değer taraması sonucunda ulaşılan oranlar kontrol edilip destek değer oranının altında kalan veri gruplarını tekrar taramaya dahil edilmeden, destek değeri üzerinde olan veri gruplarından farklı veri grupları oluşturularak veri tabanı üzerinde tekrar tarama yapılarak yeni destek değerleri hesaplanır (Eker, 2016).

Çizelge 2. 3: İkinci Destek Değer Taraması

| Ürünler | Miktar | Destek Oranı (%) |
|------------------|--------|------------------|
| Peynir, Domates | 4 | 67 |
| Peynir, Ekmek | 2 | 33 |
| Domates, Ekmek | 2 | 17 |
| Peynir, Yumurta | 1 | 33 |
| Domates, Yumurta | 1 | 17 |
| Ekmek, Yumurta | 1 | 17 |

Çizelge 2.3'deki ikinci destek değer taraması sonucunda ulaşılan oranlar tekrar kontrol edilerek değer oranının altında kalanlar dahil edilmeyecek şekilde yeni farklı gruplar oluşturulur. Bu işlem maksimum veri seti elde edilinceye kadar devam eder (Eker, 2016).

Çizelge 2. 4: Üçüncü Destek Değer Taraması

| Ürünler | Miktar | Destek Oranı(%) |
|------------------------|--------|-----------------|
| Peynir, Domates, Ekmek | 2 | 33 |

Çizelge 2.4'deki Üçüncü destek değer taraması sonuçları kontrol edildiğinde maksimum veri seti değeri ortaya çıkarıldığı görülür. Ve işlem sollandırılır.

Çizelge 2. 5: Maxsimum Kümeler ve Destek, Güven Yüzdellikleri

| ID | Geniş Kümeler | Kayıt | Destek | Güven |
|----|---------------------------------------|-------|--------|-------|
| 1 | Peynir alanlar, Domates ve Ekmek alır | 4 | %33 | %50 |
| 2 | Domates alanlar, Peynir ve Ekmek alır | 4 | %33 | %50 |
| 3 | Ekmek alanlar, Peynir ve Domates alır | 3 | %33 | %67 |
| 4 | Peynir ve Domates alanlar, Ekmek alır | 4 | %33 | %50 |
| 5 | Peynir ve Ekmek alanlar, Domates alır | 2 | %33 | %100 |

Yukarıdaki Çizelge 2.5.'de en geniş kümeye sahip olan kayıtları ve destek, güven yüzdellikleri verilmiştir. Belirtilen minumum güven %58 ve destek %27 değerlerini aşan 3 ve 5 Id'li ürün kümeleri, en geniş ürün kümeleri olmuşlardır (Eker, 2016).

3. MATERYAL VE YÖNTEM

Türk Hava Yollarının 2016 yılı içerisindeki yaz ve kış dönemine ait maskelenmiş yolcu uçuş verilerini, Apriori algoritması kullanarak ilgili verilerden kural veya kurallar ortaya çıkarılması amaçlanmıştır. Yolcu uçuş verileri, uçuş parkurları, yurt içi, yurt dışı, yolcu tipi, satış kanalı, satış tarihi ve uçuş tarihi gibi veri gruplarından oluşmaktadır.

3.1. Kullanılan Teknolojiler

3.1.1. Anaconda

Ücretsiz açık kaynak kodlu Python ve R programlama dillerinin veri bilimi ve makina öğrenmesi uygulamaları için kullanılan bir platformdur. Kullanım amacı kütüphane yönetimini ve kullanımını kolaylaştırmaktır. Uygulama kapsamında Anaconda 5.2.0 versiyonu kullanılmıştır.

3.1.1.1 Jupyter notebook

Veri temizleme, dönüştürme, nümerik simülasyon istatistiksel modelleme, veri görselleştirme ve makina öğrenme amaçlı kullanılan açık kaynak kodlu bir web uygulamasıdır. Uygulama kapsamında 5.5.0 versiyonu kullanılmıştır.

3.1.2 Python

Python, dinamik semantik ile yorumlanmış, nesne yönelimli, yüksek seviyeli bir programlama dilidir. Yüksek düzeydeki veri yapılarını hızlı bir şekilde yorumlaması uygulama geliştirmek için kullanan kişiler tarafından oldukça çekici bir hale getiriyor. Python yorumlayıcısı ve geniş standart kütüphanesi, tüm büyük platformlar için ücretsiz olarak kaynak veya ikili formda mevcuttur ve serbestçe dağıtılabilir. Python'un basit öğrenmesi kolay sözdizimi okunabilirliği vurgular ve bu nedenle program bakım maliyetini azaltır [1]. Proje kapsamında Python 3.6.6, Anaconda distribütörü ile kullanıldı.

Anaconda ücretsiz, açık kaynak kodlu, ve genellikle veri bilimi, makine öğrenmesi gibi konularda kullanılan Python ve R programlama dillerinin dağıtımını (distribütörlüğünü) yapmaktadır. Amacı kütüphane kullanımının yaygın olduğu bu dillerin paket yönetimini basitleştirmektir [2].

3.1.2.1. Numpy

Numpy, Python programla dili için yazılmış, yüksek seviye matematiksel işlemleri ve bu işlemlerin büyük ve çok boyutlu dizi veya matrisler üzerinden uygulanmasına olanak sağlayan bir açık kaynak kodlu kütüphanedir. 1995 yılında Jim Hugunin tarafından bir kaç geliştiricinin katkılarıyla Numeric adıyla geliştirildi, 2005 yılında ise Travis Oliphant Numpy'ı Numeric üzerinde değişiklik yaparak geliştirdi [3]. Proje kapsamında Numpy 1.15.1 kullanıldı.

3.1.2.2. Pandas

Pandas, Python programlama dili için veri işleme ve analizi amacıyla Wes McKinney öncülüğüyle yazılmış açık kaynak kodlu, esnek bir kütüphanedir [4]. Proje kapsamında Pandas 0.23.4 kullanıldı.

3.1.2.3. Matplotlib

Matplotlib, Python programlama dili için veri görselleştirme amacıyla ücretsiz açık kaynak kodlu kütüphanedir [5]. Proje kapsamında Matplotlib 2.2.2 kullanıldı.

3.1.2.4. Seaborn

Seaborn, matplotlib tabanlı, matplotlib'e kıyasla söz dizimi daha kolay Python programlama dili için yazılmış bir kütüphanedir [6]. Proje kapsamında Seaborn 0.9.0 kullanıldı.

3.1.2.5. Mlxtend

Mlxtend (machine learning extensions), Python programla dili için yazılmış açık kaynak kodlu ücretsiz makine öğrenmesi kütüphanesidir.

3.1.3. One-Hot encode

Tüm makine öğrenimi algoritmaları doğrudan kategorik verilerle çalışmaz. Kategorik veriler sayılara dönüştürülmelidir. One-Hot Encode, kategorik değişkenlerin ikili

vektörler olarak temsilidir. Bu ilk önce kategorik değerlerin tamsayı değerleriyle eşlenmesini gerektirir. Daha sonra, her bir tamsayı değeri, 1 ile işaretlenmiş tamsayı indeksi dışında tüm sıfır değerleri olan bir ikili vektör olarak temsil edilir.

3.1.4. Bar chart (Çubuk Grafiği)

Bar Chart, kategoriler arasında veriyi karşılaştırmak için çubukları kullanan görsel bir araçtır. Bar chart, yatay veya dikey olarak çalışabilir. Uygulamada parkurların karşılaştırılması gibi grafiklerde kullanılmıştır.

3.1.5. Heat map (Isı Haritası)

Heat Map, değerlerin renkler ile gösterildiği, matrix şeklinde, datanın iki boyutlu temsilidir [7]. Uygulamada analiz sonuçlarının matrix ve ısı grafikleri için kullanılmıştır.

3.2. Verilerin Hazırlanması

Veri tabanından, analiz de kullanılacak başlıklar belirtilerek belirli bir yıldaki belirli dönemlere ait veriler veri tabanı içerisinde alınarak excel dosyasına aktarıldı. Çekilen verilerin büyük olmasında dolayı birden fazla excel dosyasına aktarım yapıldı. Uçuş biletleme verisi, 23027214 satırdan oluşmaktadır.

Aşağıdaki Şekil 3.1.'de excel'deki analizde kullanılacak maskelenmiş uçuş biletleme verilerin bir bölümü sunulmuştur.

| TKT_SALES_Y | TKT_RTG_DOM | IT_RTGBIN_DS | TKT_NTSI_SGMR | act | ET_SGM_ACT_ARR | act | ET_SGM_ACT_DEP | M_ACT_DEP | ET_SGM_ACT_L | | |
|-------------|---------------|--------------|---------------|--------------|----------------|-----|----------------|-----------|--------------|----------|------------|
| 20160316 | International | RT | Y | TK Web | MR | IST | ISTANBUL | PRG | PRAGUE | 20160601 | 2016-06-01 |
| 20160516 | Domestic | OW | Y | Kolay Bilet | MR | ANK | ANKARA | HTY | HATAY | 20160605 | 2016-06-05 |
| 20160513 | Domestic | RT | Y | TK Mobile | MSTR | ANK | ANKARA | VAN | VAN | 20160606 | 2016-06-06 |
| 20160516 | International | RT | Y | TK Web | MS | ALA | ALMATY | IST | ISTANBUL | 20160608 | 2016-06-08 |
| 20160517 | International | RT | Y | GDS | ? | IST | ISTANBUL | GOT | GOTHENBURG | 20160607 | 2016-06-07 |
| 20160516 | Domestic | RT | Y | Websagent | MR | ERC | ERZINCAN | IST | ISTANBUL | 20160603 | 2016-06-03 |
| 20160301 | International | RT | Y | TK Web | MR | IZM | IZMIR | IST | ISTANBUL | 20160601 | 2016-06-01 |
| 20160331 | International | RT | Y | GDS | MS | ATH | ATHENS | IST | ISTANBUL | 20160604 | 2016-06-04 |
| 20160602 | Domestic | OW | Y | TK Mobile | MR | IST | ISTANBUL | ADF | ADIYAMAN | 20160603 | 2016-06-03 |
| 20160529 | Domestic | RT | Y | Websagent | MS | IZM | IZMIR | ANK | ANKARA | 20160604 | 2016-06-04 |
| 20160220 | International | RT | Y | GDS | MRS | IST | ISTANBUL | NYC | NEW YORK | 20160609 | 2016-06-09 |
| 20160531 | Domestic | RT | Y | Troya Native | MRS | ANK | ANKARA | SZF | SAMSUN | 20160608 | 2016-06-08 |
| 20160514 | International | RT | Y | TK Web | MS | DLM | DALAMAN | IST | ISTANBUL | 20160606 | 2016-06-06 |
| 20160516 | Domestic | OW | Y | Websagent | MS | DIY | DIYARBAKIR | IST | ISTANBUL | 20160606 | 2016-06-06 |
| 20160502 | International | RT | Y | GDS | MRS | PAR | PARIS | IST | ISTANBUL | 20160609 | 2016-06-09 |
| 20160530 | International | RT | Y | Troya Native | MR | IST | ISTANBUL | TBS | TBILISI | 20160601 | 2016-06-01 |
| 20160520 | Domestic | OW | Y | Websagent | MS | IST | ISTANBUL | SFQ | SANLIURFA | 20160602 | 2016-06-02 |
| 20160519 | International | RT | Y | TK Web | MS | IST | ISTANBUL | PRN | PRISTINA | 20160610 | 2016-06-10 |
| 20160719 | International | RT | Y | GDS | ? | IST | ISTANBUL | BKK | BANGKOK | 20160610 | 2016-06-10 |
| 20160531 | International | OW | Y | TK Web | MR | DNZ | DENIZLI | IST | ISTANBUL | 20160602 | 2016-06-02 |

Şekil 3.1: Maskelenmiş Uçuş Biletleme Verileri - 1

Alınan veri, hava yolu müşterilerinin hangi yılda (2016) hangi dönemlerde (yaz ve kış dönemleri) iç hat veya dış hat seferi olduğunu, tek yön veya gidiş-dönüş olduğunu, kabin tipi, Satış kanalı, müşterinin ünvanı (Bay, Bayan, Çocuk, Bebek vb.) kalkış parkuru ve varış parkuru, şehirlerin üçlü kodu, biletleme tarihi ve uçuş tarihlerini içeren bilgilerden oluşmaktadır. Şekil 3.1’de belirtilen verilerin bir bölümü sunulmuştur. Veri’deki maskelenmiş sütun adları analizlerde daha anlaşılır olması için sütun adları yeniden adlandırılmıştır. Örnek; ID_TKT_SALES_YMD sütununun gün, ay ve yıl bilgilerini içermekteydi, Bu sütun ayrı sütunlar haline dönüştürülerek YEAR ve SEASON şeklinde adlandırılmışlardır. Şekil 3.2’de dönüştürülmüş yeni adlandırılmış, veri’den bir bölüm sunulmuştur.

| YEAR | SEASON | DOM-INT | FLIGHT TYPE | CABIN CLASS | CHANNEL | PASSENGER TITLE | ARRIVAL CITY | DEPARTURE CITY | |
|---------|--------|---------|---------------|-------------|---------|-----------------|--------------|----------------|------------|
| 2036103 | 2016 | Summer | Domestic | RT | Y | TK Web | MS | ANKARA | BODRUM |
| 2036104 | 2016 | Summer | International | RT | Y | GDS | MS | ISTANBUL | DUSSELDORF |
| 2036105 | 2016 | Summer | International | RT | Y | GDS | MS | MALAGA | ISTANBUL |
| 2036106 | 2016 | Summer | Domestic | RT | Y | TK Web | MSTR | ANKARA | IZMIR |
| 2036107 | 2016 | Summer | Domestic | RT | C | TK Web | MS | ANKARA | ISTANBUL |
| 2036108 | 2016 | Summer | International | RT | Y | QuickRes | MS | ISTANBUL | BISHKEK |
| 2036109 | 2016 | Summer | International | RT | Y | GDS | MRS | LISBON | ISTANBUL |
| 2036110 | 2016 | Summer | International | RT | C | TK Web | MS | ISTANBUL | HONG KONG |
| 2036111 | 2016 | Summer | International | RT | Y | GDS | MRS | MUNICH | ISTANBUL |
| 2036112 | 2016 | Summer | Domestic | OW | Y | Troya Native | MR | ANKARA | ANTALYA |

Şekil 3.2: Maskelenmiş Uçuş Biletleme Verileri - 2

3.2.1. Veri temizleme

Excel üzerinde tutulan verilerin kontrolleri yapılmış, müşteri ünvanı, satış platformu, kabin tipi gibi veri sütunlarında veri kayıpları tespit edilmiştir. Veri kaybı tespit edilen satırlar silinerek veri analizinden çıkarılmıştır. Veriden çıkartılan satırlar sonrası toplam 2036113 satır uçuş biletleme verisi analiz edilecektir.

3.2.2. Veri entegrasyonu

Bir den fazla Excel üzerinde bulunan verilerin tek bir veri seti üzerinde analiz edilmesi için Pandas kütüphanesi kullanılarak birleştirme işlemi gerçekleştirilmiştir. Aşağıdaki Şekil 3.3.'deki ekran görüntüsünde yazılan kod ve verinin bir bölümü sunulmuştur.

```

In [48]: %%time
df1 = pd.read_excel("2016 Kış Datası.xlsx", sheet_name = 'Sheet1')

Wall time: 1min 14s

In [49]: #df1.head(10)

In [50]: %%time
df2 = pd.read_excel("2016 Yaz Datası.xlsx", sheet_name = 'Sheet1')

Wall time: 1min 18s

In [51]: #df2.head(20)

In [52]: %%time
frames = [df1,df2]
data = pd.concat(frames,ignore_index=True)

Wall time: 184 ms

In [63]: data.head(5)
Out[63]:

```

| | YEAR | SEASON | DOM-INT | FLIGHT TYPE | CABIN CLASS | CHANNEL | PASSENGER TITLE | ARRIVAL CITY | DEPARTURE CITY |
|---|------|--------|---------------|-------------|-------------|-----------|-----------------|--------------|----------------|
| 0 | 2016 | Winter | International | RT | Y | GDS | MR | ISTANBUL | SHIRAZ |
| 1 | 2016 | Winter | International | RT | Y | My ID | MR | ISTANBUL | CASABLANCA |
| 2 | 2016 | Winter | Domestic | OW | Y | QuickRes | MR | ANKARA | DIYARBAKIR |
| 3 | 2016 | Winter | International | OW | Y | Web 3.0 | MS | BARI | ISTANBUL |
| 4 | 2016 | Winter | Domestic | RT | Y | TK Mobile | MR | ANKARA | KAHRAMANMARAS |
| 5 | 2016 | Summer | International | RT | Y | GDS | MS | MALAGA | ISTANBUL |
| 6 | 2016 | Summer | Domestic | RT | Y | TK Web | MSTR | ANKARA | IZMIR |
| 7 | 2016 | Summer | Domestic | RT | C | TK Web | MS | ANKARA | ISTANBUL |
| 8 | 2016 | Summer | International | RT | Y | QuickRes | MS | ISTANBUL | BISHKEK |
| 9 | 2016 | Summer | International | RT | Y | GDS | MRS | LISBON | ISTANBUL |

Şekil 3.3: Veri Birleştirme Kodu ve Veri Seti

3.2.3. Veri seçimi

Veri setinde faydası olmayan sütunlarda tespit edilerek veri analizine dahil edilmemiştir. Örnek olarak, şehirlerin üçlü kodları analiz sürecinden çıkartılmasını diyebiliriz. Böylelikle veri boyutunda azalma olmuştur. Verilerin okunması ve analiz edilme süreçlerinde süre bakımında iyileşmeler olmuştur. Uygulamada, yaz ve kış dönemlerine ait eşit sayıda veri kullanılmıştır.

3.2.4. Veri dönüşümü

Verilerin analizi için kullanılan Apriori algoritması sadece binary (0 ve 1) türdeki girdiler ile çalıştığından dolayı, One-Hot Encode kullanılarak excel içerisindeki verilerin dönüşümü yapılmıştır.

Farklı veri tiplerinin her biri için ayrı bir sütun oluşturularak veri dönüşümü sağlanmıştır. Şekil 3.4’de dönüştürülen verilerin bir bölümü sunulmuştur.

```
In [36]: %%time
data = pd.get_dummies(data)

Wall time: 14.8 s

In [37]: data.head(20)
```

| SEASON_Summer | SEASON_Winter | DOM-INT_Domestic | DOM-INT_International | FLIGHT_TYPE_OJ | FLIGHT_TYPE_OW | FLIGHT_TYPE_RA | FLIGHT_TYPE_RS | FLIGHT_TYPE_RT | CABIN_CLASS_C | DEPARTURE_CITY_VORONEZH | DEPARTURE_CITY_WARSAW |
|---------------|---------------|------------------|-----------------------|----------------|----------------|----------------|----------------|----------------|---------------|-------------------------|-----------------------|
| 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| 0 | 1 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 1 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| 0 | 1 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |

Şekil 3.4: Verilerin Dönüştürülmesi

3.3. Verilerle İlgili Tanım ve İstatistik Bilgiler

Veri setini incelediğimiz zaman gidiş ve dönüş parkuru olmak üzere toplam 273 şehre uçuş yapıldığı görülmektedir. Bu şehirler içerisinde en fazla uçuş İstanbul’a yapıldığı en az uçuş yapılan şehirde Florence olduğu görülmektedir. Aşağıdaki Şekil 3.5’de yapılan analiz sonucu şehirlere göre uçuş sayısı verilmiştir.

| | | | | | |
|---------------|--------|---------------|------|------------------|------|
| ARRIVAL CITY | | | | | |
| ISTANBUL | 831516 | DELHI | 2840 | MUSCAT | 1195 |
| ANKARA | 195589 | HELSINKI | 2826 | VILNIUS | 1181 |
| IZMIR | 43476 | SOFIA | 2816 | DAR ES SALAAM | 1159 |
| ANTALYA | 39828 | SKOPJE | 2796 | LUXEMBOURG | 1147 |
| ADANA | 31826 | LISBON | 2709 | LEIPZIG | 1130 |
| TRABZON | 29405 | VENICE | 2669 | TEKIRDAG | 1121 |
| GAZIANTEP | 23457 | TASHKENT | 2650 | HANOI | 1115 |
| BODRUM | 21979 | MERZIFON | 2595 | SALZBURG | 1106 |
| DIYARBAKIR | 20786 | BAGHDAD | 2572 | MALE | 1091 |
| SAMSUN | 19189 | YUKSEKOVA | 2543 | BATUMI | 1084 |
| ERZURUM | 17950 | BASEL | 2532 | ZAPOROZHYE | 1076 |
| ORDU GIRESUN | 16194 | SARAJEVO | 2530 | PORTO | 1057 |
| ERCAN | 16184 | ASHGABAT | 2510 | ASTANA | 1056 |
| VAN | 16061 | CHICAGO | 2493 | LVIV | 1048 |
| DALAMAN | 15721 | ISPARTA | 2486 | TALLINN | 1030 |
| KAYSERI | 15693 | KUTAHYA | 2468 | CATANIA | 1025 |
| HATAY | 14507 | MALTA | 2465 | RIGA | 1011 |
| ELAZIG | 14320 | BISHKEK | 2446 | BUENOS AIRES | 1009 |
| MALATYA | 13328 | BEIJING | 2441 | HO CHI MINH CITY | 1001 |
| LONDON | 13178 | TIRANA | 2425 | KATHMANDU | 975 |
| SANLIURFA | 12048 | KUALA LUMPUR | 2403 | ENTEBBE | 959 |
| KONYA | 10782 | HONG KONG | 2395 | SOCHI | 939 |
| MARDIN | 10505 | SAN FRANCISCO | 2363 | TOULOUSE | 937 |
| BATMAN | 10247 | BIRMINGHAM | 2363 | BILLUND | 920 |
| PARIS | 8631 | MIAMI | 2358 | HOPA | 909 |
| KARS | 8573 | MADINAH | 2353 | DAKAR | 841 |
| DENIZLI | 8548 | GUANGZHOU | 2348 | BASRA | 836 |
| AMSTERDAM | 8391 | SHANGHAI | 2342 | ULAANBAATAR | 835 |
| DUSSELDORF | 8171 | MANILA | 2256 | DUSHANBE | 827 |
| FRANKFURT | 8119 | ODESSA | 2242 | KAZAN | 820 |
| ERZINCAN | 8034 | GOTHENBURG | 2237 | COLOMBO | 811 |
| BERLIN | 7874 | PRISTINA | 2233 | ROSTOV | 764 |
| SIVAS | 7651 | DHAKA | 2203 | DUBROVNIK | 764 |
| MUS | 7570 | WASHINGTON | 2195 | MAURITIUS | 754 |
| SIRNAK | 7562 | BOLOGNA | 2188 | SHARM EL SHEIKH | 736 |
| TEL AVIV | 7477 | LYON | 2180 | YAOUNDE | 692 |
| GAZIPASA | 7372 | NICE | 2178 | BAMAKO | 664 |
| MOSCOW | 7282 | KABUL | 2164 | EKATERINBURG | 661 |
| IGDIR | 7264 | TOKYO | 2132 | PHUKET | 657 |
| MUNICH | 6779 | LAGOS | 2108 | ASMARA | 652 |
| AGRI | 6607 | DOHA | 2103 | KINSHASA | 630 |
| ADIYAMAN | 6523 | HOUSTON | 2078 | BARI | 613 |
| VIENNA | 6450 | ALMATY | 2049 | MOMBASA | 601 |
| STUTTGART | 6356 | CASABLANCA | 2048 | SULAIMANIYAH | 598 |
| TEHRAN | 6347 | BOSTON | 1995 | YANBU | 583 |
| KAHRAMANMARAS | 6321 | ZAGREB | 1995 | DOUALA | 579 |
| DUBAI | 6150 | ISLAMABAD | 1988 | ABIDJAN | 576 |
| BRUSSELS | 6010 | CHISINAU | 1973 | MAHE ISLAND | 570 |
| NEW YORK | 5892 | THESSALONIKI | 1938 | ORAN | 565 |
| ATHENS | 5826 | BREMEN | 1935 | BOGOTA | 562 |
| BURSA | 5732 | LAHORE | 1928 | AL NAJAF | 542 |
| ZURICH | 5436 | PODGORICA | 1917 | NOUAKCHOTT | 533 |
| BAKU | 5404 | SAO PAULO | 1879 | MAZAR-I-SHARIF | 516 |
| TBILISI | 5112 | ALEXANDRIA | 1875 | CLUJ | 504 |
| BARCELONA | 5017 | KHARTOUM | 1873 | DJIBOUTI | 499 |
| HAMBURG | 5006 | JOHANNESBURG | 1867 | HAVANA | 494 |
| BEIRUT | 4883 | JAKARTA | 1832 | KIGALI | 491 |
| MILAN | 4703 | ALGIERS | 1816 | AHWAZ | 477 |
| KIEV | 4612 | KASTAMONU | 1741 | BILBAO | 472 |
| ROME | 4504 | KARACHI | 1671 | OUAGADOUGOU | 459 |
| BINGOL | 4486 | LJUBLJANA | 1648 | UFA | 444 |
| COLOGNE | 4479 | TABRIZ | 1630 | BORDEAUX | 441 |
| EDREMIT | 4218 | MALAGA | 1606 | FRIEDRICHSHAFEN | 429 |
| JEDDAH | 4138 | ACCRA | 1602 | SAMARA | 416 |
| CAIRO | 4109 | DAMMAM | 1582 | ZANZIBAR | 395 |
| BANGKOK | 4089 | ERBIL | 1572 | GRAZ | 391 |
| BUCHAREST | 4033 | MASHAD | 1556 | GYANDZHA | 391 |
| HANOVER | 3947 | NAIROBI | 1542 | NIAMEY | 383 |
| GENEVA | 3900 | SHIRAZ | 1525 | KILIMANJARO | 363 |
| COPENHAGEN | 3667 | VARNA | 1508 | COTONOU | 357 |
| BUDAPEST | 3648 | CAPE TOWN | 1498 | CONSTANTA | 346 |
| STOCKHOLM | 3556 | ST PETERSBURG | 1462 | KARLSRUHE | 345 |
| MANCHESTER | 3540 | MINSK | 1447 | GASSIM | 315 |
| PRAGUE | 3395 | VALENCIA | 1442 | VORONEZH | 314 |
| MADRID | 3300 | BAHRAIN | 1402 | CONSTANTINE | 313 |
| RIYADH | 3272 | ABUJA | 1393 | CARACAS | 307 |
| KUWAIT | 3191 | KHERSON | 1388 | MUNSTER | 306 |
| NUREMBERG | 3174 | EDINBURGH | 1374 | DURBAN | 296 |
| AMMAN | 3137 | MARSEILLE | 1369 | KOSICE | 291 |
| NEVSEHIR | 3097 | ADDIS ABABA | 1369 | TURIN | 285 |
| TUNIS | 3030 | TAIPEI | 1351 | ANTANANARIVO | 266 |
| TORONTO | 3003 | ABU DHABI | 1312 | TAIF | 260 |
| BELGRADE | 3002 | MOGADISHU | 1310 | LIBREVILLE | 256 |
| LOS ANGELES | 2994 | MONTREAL | 1289 | CONAKRY | 233 |
| CANAKKALE | 2973 | KHARKOV | 1283 | STAVROPOL | 228 |
| OSLO | 2961 | NAPLES | 1274 | PANAMA CITY | 211 |
| MUMBAI | 2950 | WARSAW | 1255 | MAPUTO | 207 |
| DUBLIN | 2942 | ATLANTA | 1232 | NAKHICHEVAN | 202 |
| SEOUL | 2926 | HURGHADA | 1228 | FISA | 201 |
| SINOP | 2876 | KOCAELI | 1213 | NDJAMENA | 145 |
| SINGAPORE | 2844 | ISFAHAN | 1196 | FLORENCE | 70 |

Şekil 3.5: Şehirlere Göre Uçuş Sayısı

Veri setimizde bulunan satış kanallarının açılımları (tanım) aşağıdaki Şekil 3.6'da verilmiştir.

| Satış Kanalı | Tanım |
|---------------------|--|
| GDS | Acenta |
| Troya Native | Satış Ofisi |
| Websagent | Online Acentalar |
| TK Web | THY Web Sitesi |
| TK Mobile | THy Mobil Uygulama |
| QuickRes | Çağrı Merkezi |
| Web 3.0 | THY Web Sitesi |
| AJet Web | Anadolujet Web |
| Kolay Bilet | Kolay Bilet (Acenta) |
| Web Pass | Pass bilet (Personel) |
| AJet Mobile | Anadolujet Mobil Uygulama |
| My ID | Anlaşmalı Personel Bileti (Diğer Havayolu Personeli) |
| CrewFly | Anlaşmalı Personel Bileti (Diğer Havayolu Personeli) |
| TCC Online | Kurumsal Site Satış (Kurumsal Anlaşmalı Fiyatlar) |
| TCC Convention | Kurumsal Site Satış (Satış Ofisi) |
| Quick Check-in | Checkin Sistemi |

Şekil 3.6: Satış Kanalları

Veri setimizdeki satış kanallarını incelediğimiz zaman en fazla biletleme yapılan kanalın GDS olduğunu en az biletleme yapılan satış kanalının ise Quick Check-in olduğu görülmüştür. Aşağıdaki Şekil 3.7'de satış kanalları üzerinden yapılan biletleme sayıları verilmiştir.

```

In [103]: data.groupby("CHANNEL")["CHANNEL"].count().sort_values(ascending=False)
Out[103]: CHANNEL
GDS                536497
Troya Native       264967
Websagent          221263
TK Web             211334
TK Mobile          209906
QuickRes           203958
Web 3.0            159467
AJet Web           90659
Kolay Bilet        62135
Web Pass           36065
AJet Mobile        25296
My ID              10410
Other              1473
CrewFly            1322
TCC Online         1236
TCC Convention     122
Quick Check-in     3
Name: CHANNEL, dtype: int64

```

Şekil 3.7: Satış Kanalları Üzerinden Yapılan Biletleme Sayısı

Veri setimizdeki yolcu ünvanlarının karşılıkları (tanımları) aşağıdaki Şekil 3.8'de sunulmuştur.

| Yolcu Tipi | Tanım |
|------------|------------------------------|
| MR | Erkek Yolcu |
| MS | Kadın Yolcu (Evli) |
| MRS | Kadın Yolcu |
| MSTR | Erkek Çocuk veya Bebek Yolcu |
| MISS | Kız Çocuk veya Bebek Yolcu |
| CHD | Çocuk Yolcu (2 - 11) |
| DR | Doktor (Ünvanı) Yolcu |
| INF | Bebek Yolcu (2 yaş altı) |
| PROF | Profesör (Ünvanı) Yolcu |
| MRST | Resmi (Ünvanı) Yolcu |

Şekil 3.8: Yolcu Ünvanlarının Tanımları

Veri setinde bulunan yolcu ünvanlarına göre yurt içi (Domestic) ve yurt dışı (International) dağılımları aşağıdaki Şekil 3.9'da verilmiştir. Bu analiz sonuçlarına göre en fazla erkek yolcular (MR) yurt dışı (International) biletleme yaptığını en az ise resmi yolcu (MRST) tipinin yurt için (Domestic) yaptığı bilgisine ulaşılmıştır.

```
In [76]: data.groupby(['PASSENGER TITLE', 'DOM-INT'])['DOM-INT'].count()

Out[76]: PASSENGER TITLE  DOM-INT
CHD                    Domestic      7
                    International  13906
DR                     Domestic    140
                    International  2458
INF                    Domestic     1
                    International   752
MISS                   Domestic  36779
                    International  16538
MR                     Domestic  403363
                    International  535409
MRS                    Domestic  104883
                    International  241864
MRST                   Domestic     1
                    International   3
MS                     Domestic  313941
                    International  263752
MSTR                   Domestic   71886
                    International  30105
PROF                   Domestic    11
                    International   314
Name: DOM-INT, dtype: int64
```

Şekil 3.9: Yolcu Ünvanları Göre Yurt İçi ve Yurt Dışı Biletleme Sayıları

Veri setimiz içerisinde yer alan kabin tipi ve uçuş tipi sütunlarını analiz ettiğimizde, Y (Economy) ve C (Business) olmak üzere iki kabin tipi olarak biletleme yapıldığı görülmüştür. 2016 yılı yaz ve kış dönemlerine ait uçuşlarda 1906865 kişi kabin tipi Y (Economy) olarak, 129248 kişi ise kabin tipi C (Business) olarak biletleme yapıldığı görülmektedir.

```
In [108]: data.groupby("CABIN CLASS")["CABIN CLASS"].count().sort_values(ascending=False)

Out[108]: CABIN CLASS
Y      1906865
C      129248
Name: CABIN CLASS, dtype: int64
```

Şekil 3.10: Kabin Tipleri

Uçuş tipi 5 (beş) farklı tip olarak biletleme yapıldığı analiz sonucu ortaya çıkmıştır. Bu tiplerin tanımları aşağıdaki gibidir,

- RT : Gidiş – Dönüş (İstanbul → Ankara – Ankara → İstanbul)
- OW: Tek Yön (İstanbul → Londra)
- RS : Farklı Hava Yollarına Ait İkili Kod

- RA : Farklı Hava Yollarına Ait İkili Kod
- OJ : Gidiş – Dönüş farklı parkurlar (İstanbul → Ankara Gidiş Ankara → İzmir Dönüş)

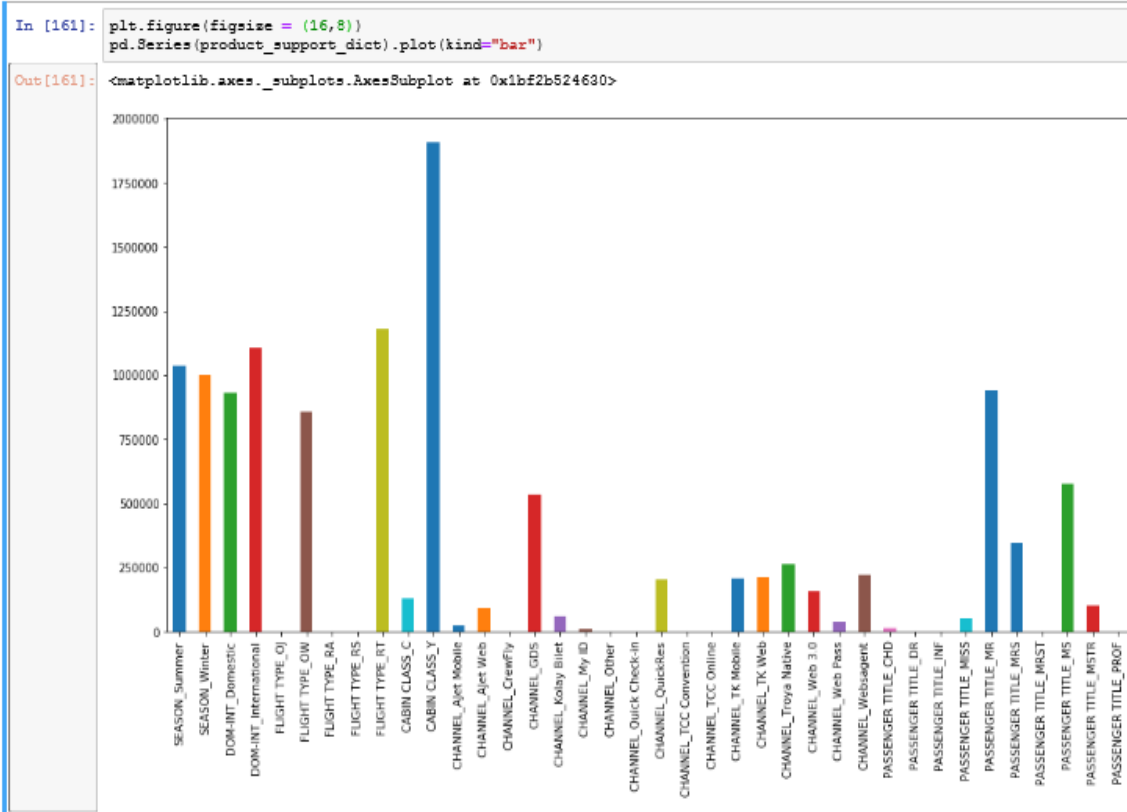
Şekil 3.11’de veri setindeki uçuş tipi analizi sonucu en fazla RT (Gidiş-Dönüş) 1178423 en az ise RS (Farklı hava yolu kodu) 309 adet biletleme yapıldığı görülmektedir.

```
In [109]: data.groupby("FLIGHT TYPE")["FLIGHT TYPE"].count().sort_values(ascending=False)

Out[109]: FLIGHT TYPE
RT      1178423
OW      855009
RA       1451
OJ        921
RS         309
Name: FLIGHT TYPE, dtype: int64
```

Şekil 3.11: Uçuş Tipi Analizi Sonucu

Veri setinde bulunan tüm alanların grafiksel (bar) dökümü Şekil 3.12’de sunulmuştur.



Şekil 3.12: Veri Seti Analiz Grafiği

3.4. 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, mininum destek ($\text{min_support}=0.1$) değeri 0.1, lift ($\text{min_threshold}=1$) değeri 1 verilerek analiz edilmiştir. Belirtilen support değerini aşan 205 farklı grup ortaya çıkarılmıştır. Şekil 3.14’de destek değeri sonrası ortaya çıkartılan örüntülerin bir kısmı sunulmuştur. 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 kontrol 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 [8]. Aşağıdaki Şekil 3.13’de uygulamada kullanılan paket kodun bir kısmı sunulmuştur.

```

In [1]: def apriori(df, min_support=0.5, use_colnames=False, max_len=None, n_jobs=1):
    allowed_val = {0, 1, True, False}
    unique_val = np.unique(df.values.ravel())
    for val in unique_val:
        if val not in allowed_val:
            s = ('The allowed values for a DataFrame'
                ' are True, False, 0, 1. Found value %s' % (val))
            raise ValueError(s)
    is_sparse = hasattr(df, "to_coo")
    if is_sparse:
        X = df.to_coo().tocsc()
        support = np.array(np.sum(X, axis=0) / float(X.shape[0])).reshape(-1)
    else:
        X = df.values
        support = (np.sum(X, axis=0) / float(X.shape[0]))
    ary_col_idx = np.arange(X.shape[1])
    support_dict = {1: support[support >= min_support]}
    itemset_dict = {1: ary_col_idx[support >= min_support].reshape(-1, 1)}
    max_itemset = 1
    rows_count = float(X.shape[0])

    if max_len is None:
        max_len = float('inf')

    while max_itemset and max_itemset < max_len:
        next_max_itemset = max_itemset + 1
        combin = generate_new_combinations(itemset_dict[max_itemset])
        frequent_items = []
        frequent_items_support = []

        if is_sparse:
            all_ones = np.ones((X.shape[0], next_max_itemset))
        for c in combin:
            if is_sparse:
                together = np.all(X[:, c] == all_ones, axis=1)
            else:
                together = X[:, c].all(axis=1)
            support = together.sum() / rows_count
            if support >= min_support:
                frequent_items.append(c)
                frequent_items_support.append(support)

        if frequent_items:
            itemset_dict[next_max_itemset] = np.array(frequent_items)
            support_dict[next_max_itemset] = np.array(frequent_items_support)
            max_itemset = next_max_itemset
        else:
            max_itemset = 0

    all_res = []
    for k in sorted(itemset_dict):
        support = pd.Series(support_dict[k])
        itemsets = pd.Series([frozenset(i) for i in itemset_dict[k]])

```

Şekil 3.13: Uygulamadaki Apriori Kodu [8]

```
In [165]: from mlxtend.frequent_patterns import apriori
         from mlxtend.frequent_patterns import association_rules
```

```
In [170]: frequent_itemsets = apriori(data,min_support=0.1,use_colnames=True)
         frequent_itemsets
```

Out[170]:

| | support | itemsets |
|-----|----------|---|
| 0 | 0.508603 | (SEASON_Summer) |
| 1 | 0.491397 | (SEASON_Winter) |
| 2 | 0.457250 | (DOM-INT_Domestic) |
| 3 | 0.542750 | (DOM-INT_International) |
| 4 | 0.419922 | (FLIGHT_TYPE_OW) |
| 5 | 0.578761 | (FLIGHT_TYPE_RT) |
| 6 | 0.936522 | (CABIN_CLASS_Y) |
| 7 | 0.263491 | (CHANNEL_GDS) |
| 8 | 0.100170 | (CHANNEL_QuickRes) |
| 9 | 0.103092 | (CHANNEL_TK Mobile) |
| 10 | 0.103793 | (CHANNEL_TK Web) |
| 11 | 0.130134 | (CHANNEL_Troya Native) |
| 12 | 0.108669 | (CHANNEL_Websagent) |
| 13 | 0.461061 | (PASSENGER_TITLE_MR) |
| 14 | 0.170299 | (PASSENGER_TITLE_MRS) |
| 15 | 0.283723 | (PASSENGER_TITLE_MS) |
| 16 | 0.408384 | (ARRIVAL_CITY_ISTANBUL) |
| 17 | 0.422582 | (DEPARTURE_CITY_ISTANBUL) |
| 18 | 0.200911 | (DOM-INT_Domestic, SEASON_Summer) |
| 19 | 0.307692 | (SEASON_Summer, DOM-INT_International) |
| 20 | 0.188172 | (SEASON_Summer, FLIGHT_TYPE_OW) |
| ... | ... | ... |
| 181 | 0.123219 | (DOM-INT_Domestic, PASSENGER_TITLE_MR, CABIN C... |
| 182 | 0.139984 | (CABIN_CLASS_Y, DOM-INT_International, FLIGHT ... |
| 183 | 0.118888 | (PASSENGER_TITLE_MR, DOM-INT_International, CA... |
| 184 | 0.102332 | (ARRIVAL_CITY_ISTANBUL, DOM-INT_International,... |
| 185 | 0.104092 | (SEASON_Winter, DOM-INT_International, CABIN C... |
| 186 | 0.115204 | (PASSENGER_TITLE_MR, FLIGHT_TYPE_OW, CABIN CLA... |
| 187 | 0.126633 | (CABIN_CLASS_Y, PASSENGER_TITLE_MR, FLIGHT TYP... |
| 188 | 0.101092 | (CABIN_CLASS_Y, ARRIVAL_CITY_ISTANBUL, FLIGHT ... |
| 189 | 0.102746 | (DEPARTURE_CITY_ISTANBUL, CABIN_CLASS_Y, FLIGH... |
| 190 | 0.109548 | (DOM-INT_Domestic, PASSENGER_TITLE_MR, CABIN C... |
| 191 | 0.169797 | (CHANNEL_GDS, CABIN_CLASS_Y, FLIGHT_TYPE_RT, D... |
| 192 | 0.157529 | (CABIN_CLASS_Y, PASSENGER_TITLE_MR, FLIGHT TYP... |
| 193 | 0.174638 | (CABIN_CLASS_Y, ARRIVAL_CITY_ISTANBUL, FLIGHT ... |
| 194 | 0.176058 | (CABIN_CLASS_Y, DOM-INT_International, FLIGHT ... |
| 195 | 0.107180 | (CHANNEL_GDS, PASSENGER_TITLE_MR, CABIN_CLASS_... |
| 196 | 0.114111 | (CHANNEL_GDS, ARRIVAL_CITY_ISTANBUL, CABIN CLA... |
| 197 | 0.115613 | (CHANNEL_GDS, DOM-INT_International, CABIN CLA... |
| 198 | 0.111164 | (ARRIVAL_CITY_ISTANBUL, PASSENGER_TITLE_MR, CA... |
| 199 | 0.113490 | (PASSENGER_TITLE_MR, DOM-INT_International, CA... |
| 200 | 0.102276 | (CABIN_CLASS_Y, PASSENGER_TITLE_MR, ARRIVAL CI... |
| 201 | 0.104077 | (CABIN_CLASS_Y, PASSENGER_TITLE_MR, FLIGHT TYP... |
| 202 | 0.107702 | (FLIGHT_TYPE_RT, DOM-INT_International, SEASON... |
| 203 | 0.106184 | (FLIGHT_TYPE_RT, DOM-INT_International, SEASON... |
| 204 | 0.106894 | (FLIGHT_TYPE_RT, DOM-INT_International, SEASON... |

205 rows x 2 columns

Şekil 3.14: Support Değerini Aşan Gruplar

Veri setinin analizi sonucu toplamda 824 adet yorumlanmak üzere kural ortaya çıkarılmıştır. Ortaya çıkarılan kuralların bir bölümü Şekil 3.15’de sunulmuştur.

```
In [173]: rules = association_rules(frequent_itemsets, metric="lift", min_threshold=1)
rules
```

```
Out[173]:
```

| | antecedents | consequents | antecedent support | consequent support | support | confidence | lift | leverage | conviction |
|-----|--|---|--------------------|--------------------|----------|------------|----------|----------|------------|
| 0 | (SEASON_Summer) | (DOM-INT_International) | 0.508803 | 0.542750 | 0.307892 | 0.604974 | 1.114646 | 0.031647 | 1.157519 |
| 1 | (DOM-INT_International) | (SEASON_Summer) | 0.542750 | 0.508803 | 0.307892 | 0.566912 | 1.114646 | 0.031647 | 1.134636 |
| 2 | (SEASON_Summer) | (FLIGHT TYPE_RT) | 0.508803 | 0.578761 | 0.319535 | 0.628261 | 1.085527 | 0.025176 | 1.133157 |
| 3 | (FLIGHT TYPE_RT) | (SEASON_Summer) | 0.578761 | 0.508803 | 0.319535 | 0.552102 | 1.085527 | 0.025176 | 1.097119 |
| 4 | (SEASON_Summer) | (CABIN CLASS_Y) | 0.508803 | 0.936522 | 0.476691 | 0.937255 | 1.000783 | 0.000373 | 1.011680 |
| 5 | (CABIN CLASS_Y) | (SEASON_Summer) | 0.936522 | 0.508803 | 0.476691 | 0.509001 | 1.000783 | 0.000373 | 1.000811 |
| 6 | (CHANNEL_GDS) | (SEASON_Summer) | 0.263491 | 0.508803 | 0.155608 | 0.590562 | 1.161146 | 0.021596 | 1.200176 |
| 7 | (SEASON_Summer) | (CHANNEL_GDS) | 0.508803 | 0.263491 | 0.155608 | 0.305951 | 1.161146 | 0.021596 | 1.061178 |
| 8 | (SEASON_Summer) | (PASSENGER TITLE_MRS) | 0.508803 | 0.170299 | 0.103332 | 0.203169 | 1.193015 | 0.016718 | 1.041251 |
| 9 | (PASSENGER TITLE_MRS) | (SEASON_Summer) | 0.170299 | 0.508803 | 0.103332 | 0.606771 | 1.193015 | 0.016718 | 1.249646 |
| 10 | (PASSENGER TITLE_MS) | (SEASON_Summer) | 0.283723 | 0.508803 | 0.163791 | 0.577291 | 1.135053 | 0.019488 | 1.162495 |
| 11 | (SEASON_Summer) | (PASSENGER TITLE_MS) | 0.508803 | 0.283723 | 0.163791 | 0.322041 | 1.135053 | 0.019488 | 1.055519 |
| 12 | (ARRIVAL CITY_ISTANBUL) | (SEASON_Summer) | 0.408384 | 0.508803 | 0.211697 | 0.518377 | 1.019218 | 0.003992 | 1.020295 |
| 13 | (SEASON_Summer) | (ARRIVAL CITY_ISTANBUL) | 0.508803 | 0.408384 | 0.211697 | 0.416232 | 1.019218 | 0.003992 | 1.013444 |
| 14 | (SEASON_Summer) | (DEPARTURE CITY_ISTANBUL) | 0.508803 | 0.422582 | 0.221205 | 0.434927 | 1.029214 | 0.006279 | 1.021847 |
| 15 | (DEPARTURE CITY_ISTANBUL) | (SEASON_Summer) | 0.422582 | 0.508803 | 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-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... | 0.307892 | 0.238804 | 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 | 1.090967 |
| 816 | (DEPARTURE CITY_ISTANBUL, DOM-INT_International) | (CABIN CLASS_Y, SEASON_Summer, FLIGHT TYPE_RT) | 0.266996 | 0.298100 | 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... | 0.390246 | 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... | 0.508803 | 0.170058 | 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 |

824 rows x 9 columns

Şekil 3.15: Oluşan Kurallar

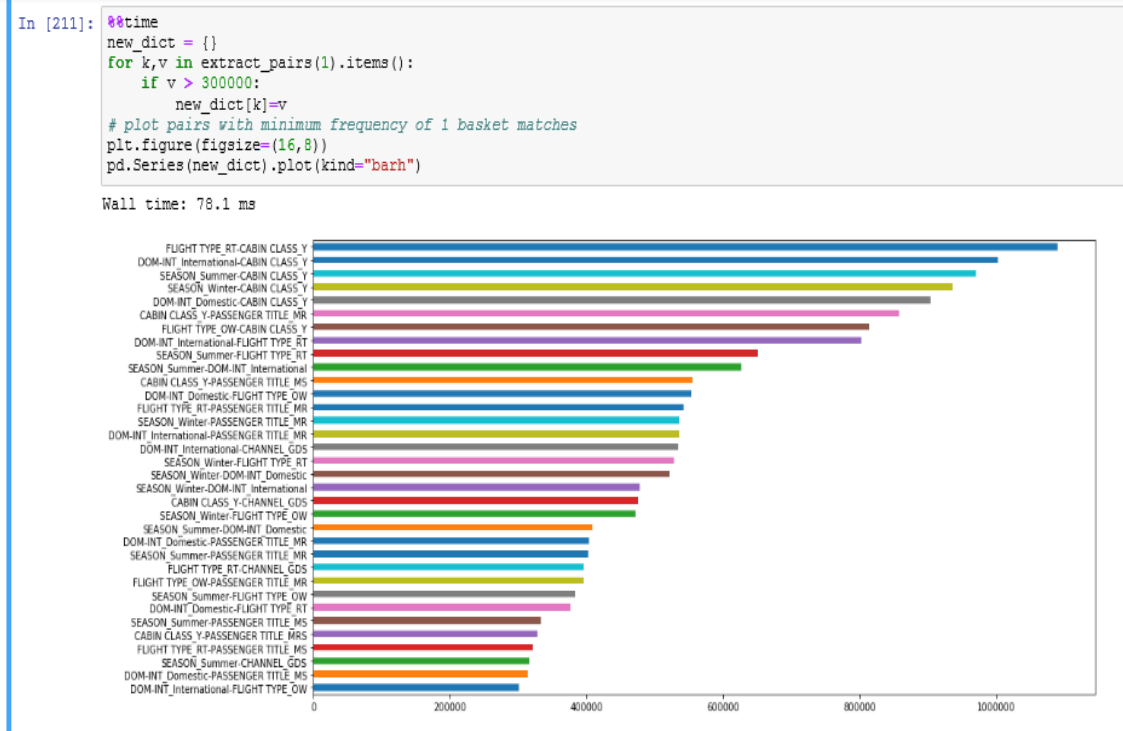
Uygulama sonucu elde edilen sonuçlarda öne çıkan kurallar, kış döneminde erkek yolcu tipinin iç hat uçuşlar için ekonomi sınıfını %96 güven, %12 destek değeri ile ekonomi sınıfı tercih ettiği Ayrıca 1,13 lift değeri 1'den büyük olması nedeniyle bağımsız olduğunu ortaya çıkarmıştır. Yine bir başka öne 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.

Yapılan analizler sonucu veri görselleştirmesi metotlarından da faydalanıldı. Birliktelik kuralı veri görselleştirmelerinde en çok kullanılan grafikler ısı haritası ve bar chart'dır. Bu iki görselleştirme metodu da data üzerinden kullanıldı. Şekil 3.16'de kuralların ısı haritasına bakıldığında açık renkli ikili ürün gruplarının beraber alınma sıklığının daha fazla olduğu, koyu renkli ürün gruplarının alınma sıklığının daha az olduğu görülmektedir. Dolayısıyla bu durumun kurallara da yansımaları beklenmektedir. Örneğin, yurt dışı biletleme yapan yolcuların büyük bir bölümü ekonomi sınıfında yolculuk yapmıştır.



Şekil 3.16: Isı Haritası

Isı haritasının yorumlanmasının güç olduğu durumlarda, bar chart'ta sadece ilişkisi olan ürün grupları görüldüğü için tercih edilebiliyor. Bar chart'ın incelenmesi sonucu erkek yolcuların kadın yolculara kıyasla daha fazla ekonomi sınıfı ile yolculuk yaptığı gözlemlenmiştir. Aşağıdaki Şekil 3.17'da ise algoritma sonrası öne çıkan örüntüleri çubuk grafiği sunulmuştur.



Şekil 3.17: Çubuk Grafiği

4. SONUÇ VE ÖNERİLER

Bir çok kurum ve kuruluşta olduğu gibi hava yolu şirketleri de veri madenciliği yöntemleri açısından büyük potansiyel barındırmaktadır. Kritik öneme sahip müşteriye direkt dokunan stratejik kararlarda, veri madenciliği yöntemleri etkin kullanılmaktadır. Gelecekte bu yöntemlerin kullanım sıklığının daha da artması kaçınılmazdır.

Yapılan analiz sonrası hava yolu müşterilerinin büyük pazarlama fırsatlarına açık olduğu öngörülmektedir. Yapılan analiz sonucunda, hava yolu müşterilerinin biletleme davranışlarını analiz edilerek, onlara daha iyi hizmet verebilmek ve hava yolu şirketlerine, ellerine bulundurdukları veriye, veri madenciliği uygulandıktan sonra çıkan kurallar ve sonuçlara göre müşterilerinin biletleme davranışlarına göre aksyonlar alarak daha iyi müşteri deneyimi sağlayabilecekleri öneriler bulunulmuştur.

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 seyahatlerini 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 yapılabilir. 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 aşmadığı 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.

KAYNAKLAR

- Agrawal, R., Srikant, R.** (1993). Fast Algorithms For Mining Association Rules, Conference on Very Large Databases, Santiago, Chile. 487-499.
- Akpınar, H.** (2014). Data: Veri Madenciliği Veri Analizi, Papatya Yayıncılık Eğitim.
- 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.
- Argüden, Y. ve Erşahin, B.** (2010). Veri Madenciliği. ISBN: 978-975-93641-9-9
- Berry, M. J. A. ve Linoff, G. S.** (2004). Data Mining Techniques for Marketing, Sales and Customer Relationship Management, Wiley Publishing, Inc., Indianapolis.
- Dalkılıç, F., Aydın, O.** (2017). Dokuz Eylül Üniversitesi İktisadi ve İdari Bilimler Fakültesi Öğrencilerinin Devamsızlık Davranışlarını Etkileyen Faktörler. Journal of Higher Education and Science, vol.7, No:3, s.546.
- Döşlü A.** (2008). Veri Madenciliğinde Market Sepet Analizi ve Birliktelik Kurallarının Belirlenmesi, Yüksek Lisans Tezi, Yıldız Teknik Üniversitesi, Fen Bilimleri Enstitüsü.
- Durdu, M.** (2012). Müşteri İlişkileri Yönetiminde Veri Madenciliği Uygulaması: Bir Perakende Mağazasında Market Sepet Analizi, Yüksek Lisans Tezi, Dokuz Eylül 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.

- Ekim, U.** (2011). Veri Madenciliği Algoritmalarını Kullanarak Öğrenci Verilerinden Birliktelik Kurallarının Çıkarılması, Yüksek Lisans Tezi, Selçuk 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 Madenciliğinde Apriori Algoritması ve Apriori Algoritmasının Farklı Veri Kümelerinde Uygulanması, Yüksek Lisans Tezi, Trakya Üniversitesi, Fen Bilimleri Enstitüsü.
- Gündüz, D.** (2015). İş zekası Uygulamaları ve Pazar Sepeti Analizi, Yüksek Lisans Tezi, Trakya Üniversitesi, Fen Bilimleri Enstitüsü.
- Gündüz, S.** (2017). Veri Madenciliğinde Kullanılan Birliktelik Analizi ve Merkez Sepet Analizi: Bir Uygulama, Yüksek Lisans Tezi, Ondokuz Mayıs Üniversitesi, Fen Bilimleri Enstitüsü.
- Han, J., Pei, J., Kamber, M.** (2001). Data Mining Concepts And Techniques, Eker, M. E Elsevier.
- Jacobs, P.** (1999). Data Mining: What General Managers Need to Know, Harvard Management Update 4, 8.
- Jain, R.** (2017). Application Of Data Mining Techniques For Predicting Students Academic Performance: A Study, International Journal of Innovation in Engineering Research & Management, Vol. 04, No.3, June 2017.
- Jie, X., Wang, H., Fei, M., Du, D., Güneş, Q., Yang, T.C.** (2018). Anomaly Behavior Detection and Reliability Assessment of Control Systems Based on Association Rules, International Journal of Critical Infrastructure Protection, Volume 22, Pages 90-99
- Karaibrahimoğlu, A., Genç A.** (2014). Meme Kanseri Verisinde Apriori Algoritması ile Kural Çıkarma, Selçuk Tıp Dergisi, Cilt: 30, Sayı: 3.
- 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
- Malık Malık, Z. M.** (2018). Gözetimsiz Makine Öğrenim Teknikleri ile Miktara Dayalı Negatif Birliktelik Kural Madenciliği, Yüksek Lisans Tezi, Türk Hava Kurumu Üniversitesi, Fen Bilimleri Enstitüsü.
- Onat, A.** (2008). Veri Madenciliğinin Web Tabanlı Uygulamalarda İnsan Uyumluluklarının Tesbiti Üzerine Bir Çalışma, Yüksek Lisans Tezi, Selçuk Üniversitesi, Fen Bilimleri Enstitüsü.
- 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 Belirlenmesinde Veri Madenciliği Uygulaması, Yüksek Lisans Tezi, Marmara Üniversitesi, Fen Bilimleri Enstitüsü.
- Özkan, Y.** (2008). Veri Madenciliği Yöntemleri, Papatya Yayıncılık Eğitim.
- Öztemiz, F.** (2017). Apriori Algoritması ile Müşteri Bazlı Market Sepet Analizi ve Ürün Satış Tahmini, Yüksek Lisans Tezi, İnönü Üniversitesi, Fen Bilimleri Enstitüsü.
- Roiger, R. ve Geatz, M.** (2003). *Data Mining: A Tutorial-Based Primer*, Boston MA: Addison Wesley.
- Sadıqmal F.** (2015). Implementation of Some Medical Data Using Apriori Algorithm, Master Thesis, Sakarya University, Institute Of Science And Technology.
- Şen, F.** (2008). Veri Madenciliği ile Birliktelik Kurallarının Bulunması, Yüksek Lisans Tezi, Sakarya Üniversitesi, Fen Bilimleri Enstitüsü.
- Silahtaroglu, G** (2014). Veri Madenciliği. ISBN: 978-975-6797-81-5
- Sıramkaya, E.** (2005). Veri Madenciliğinde Bulanık Mantık Uygulaması, Yüksek Lisans Tezi, Selçuk Üniversitesi, Fen Bilimleri Enstitüsü.
- Toprak, S.** (2004). Data Mining For Rule Discovery in Relational Databases, Master Thesis, Middle East Technical University, Computer Engineering.
- Toprak, S. D.** (2005). A New Hybrid Multi-Relational Data Mining Technique, Master Thesis, Middle East Technical University, Computer Engineering.

- 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 Institute 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.

İnternet Kaynakları:

- [1] < <https://www.python.org/doc/essays/blurb/> >, alındığı tarih: 21.11.2018.
- [2] < <https://www.anaconda.com/what-is-anaconda>>, alındığı tarih: 27.10.2018.
- [3] < <http://www.numpy.org> > , alındığı tarih: 27.10.2018.
- [4] < https://www.dlr.de/sc/Portaldata/15/Resources/dokumente/pyhpc2011/submissions/pyhpc2011_submission_9.pdf>, alındığı tarih: 27.10.2018.
- [5] < <https://matplotlib.org>>, alındığı tarih: 27.10.2018.
- [6] < <https://seaborn.pydata.org>>, alındığı tarih: 27.10.2018.
- [7] < <https://searchbusinessanalytics.techtarget.com/definition/heat-map>>, alındığı tarih: 27.10.2018.
- [8] < <https://github.com/rasbt/mlxtend> >, Alındığı tarih: 27.11.2018

ÖZGEÇMİŞ



Adı-Soyad : Muhammed Metin ULUYARDIMCI

Doğum Tarihi ve Yeri : 05.09.1990, Şanlıurfa

E-posta : muluyardimci@gmail.com

ÖĞRENİM DURUMU:

- **Lisans** : 2012 - 2015, İstanbul Arel Üniversitesi, Mühendislik ve Mimarlık Fakültesi, Bilgisayar Mühendisliği
- **Yüksek Lisans** : 2016 – 2018, İstanbul Aydın Üniversitesi, Fen Bilimleri Enstitüsü, Bilgisayar Mühendisliği Anabilim Dalı