T.C
İNÖNÜ ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

APRIORI ALGORİTMASI İLE MÜŞTERİ BAZLI MARKET SEPET ANALİZİ VE ÜRÜN SATIŞ TAHMİNİ

FURKAN ÖZTEMİZ

YÜKSEK LİSANS TEZİ

BİLGİSAYAR MÜHENDİSLİĞİ ANABİLİM DALI

TEMMUZ-2017

T.C
İNÖNÜ ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

APRIORI ALGORİTMASI İLE MÜŞTERİ BAZLI MARKET SEPET ANALİZİ VE ÜRÜN SATIŞ TAHMİNİ

FURKAN ÖZTEMİZ

YÜKSEK LİSANS TEZİ

BİLGİSAYAR MÜHENDİSLİĞİ ANABİLİM DALI

TEMMUZ 2017

Tezin Başlığı: APRIORI ALGORİTMASI İLE MÜŞTERİ BAZLI MARKET SEPET ANALİZİ VE ÜRÜN SATIŞ TAHMİNİ

Tezi Hazırlayan: Furkan ÖZTEMİZ

Sınav Tarihi: 31 Temmuz 2017

Yukarıda adı geçen tez jürimizce değerlendirilerek Bilgisayar Mühendisliği Ana Bilim Dalında Yüksek Lisans Tezi olarak kabul edilmiştir.

Sınav Jüri Üyeleri

Tez Danışmanı : **Prof. Dr. Serdar Ethem HAMAMCI**

İnönü Üniversitesi

Prof. Dr. Ali KARCI

İnönü Üniversitesi

Doç. Dr. Resul DAŞ

Fırat Üniversitesi Yazılım Mühendisliği Bölümü

Prof. Dr. Halil İbrahim ADIGÜZEL

Enstitü Müdürü

ONUR SÖZÜ

Yüksek Lisans Tezi olarak sunduğum "APRIORI ALGORİTMASI İLE MÜŞTERİ BAZLI MARKET SEPET ANALİZİ VE ÜRÜN SATIŞ TAHMİNİ" başlıklı bu çalışmanın bilimsel ahlak ve geleneklere aykırı düşecek bir yardıma başvurmaksızın tarafımdan yazıldığını ve yararlandığım bütün kaynakların, hem metin içinde hem de kaynakçada yönetimine uygun biçimde gösterilenlerden oluştuğunu belirtir, bunu onurumla doğrularım.

Furkan ÖZTEMİZ

ÖZET

Yüksek Lisans Tezi

APRIORI ALGORİTMASI İLE MÜŞTERİ BAZLI MARKET SEPET ANALİZİ VE ÜRÜN SATIŞ TAHMİNİ

Furkan ÖZTEMİZ

İnönü Üniversitesi

Fen Bilimleri Enstitüsü

Bilgisayar Mühendisliği Anabilim Dalı

88 + x

2017

Danışman: Prof. Dr. Serdar Ethem Hamamcı

Veri madenciliğinin yaygın kullanımı ve gelişmesi ile büyük veri yığınları içerisinden anlamlı örüntüler çıkarmayı mümkün kılmıştır. Birçok sektör veri madenciliği yöntemlerini kendi verileri üzerinde aktif olarak kullanmıştır. Bu tez çalışmasında apriori algoritması kullanılarak perakende sektöründe hizmet veren özel bir firmaya ait yüksek miktardaki satış verilerinin analizi yapılmıştır. Analiz işleminin temel amacı müşteri memnuniyetini sağlamak, firmaya olan bağlılığı artırmak, firma satış grafiğini yükseltmektir. Müşterilerin satın aldıkları ürünler arasındaki birliktelik kurallarının tespit edilmesi ile bu kuralların temsil ettikleri müşteri kitlelerinin belirlenmesi sağlanmıştır. Böylece firmanın tek bir kampanya yaparak müşterileri genellemesi yerine çok sayıda kampanya oluşturarak az sayıda ürünü doğru müşterilere sunması sağlanmıştır. Bu özel kampanyalar ile müşterilerin kendilerini daha özel hissetmeleri ve firmaya olan bağlılıklarının artırılması hedeflenmiştir. Birliktelik kuralı yöntemleri uygulanarak elde edilen analiz sonuçları neticesinde müşteriler için hazırlanan kampanyalarda sunulan ürünlere ait satış miktarı tahminleri yapay sinir ağları yöntemi ile yapılmıştır. Bu sayede kampanyanın müşteriler tarafından ne kadar ilgi göreceği hakkında önemli bilgiler elde edinilmiştir. Bu analiz sürecinde verilerin sistemli bir şekilde kayıt altına alınması ve analize uygun hale getirilmesi işlemi MSSQL Server veri tabanı sistemi üzerinden yapılmıştır. Birliktelik kurallarının oluşturulması için R programlama dili ve R Studio geliştirme ortamı kullanılmıştır. Yöntem olarak oldukça popüler olan apriori algoritması tercih edilmiştir. YSA kullanılarak gerçekleştirilen tahmin işleminde ise MATLAB kullanılmıştır. Elde edilen birliktelik kurallarını temsil eden müşterilerin tespit edilmesinde gösterge panelleri tasarlanarak sonuçların daha kolay ve daha anlaşılır bir yapıya kavuşması sağlanmıştır.

ANAHTAR KELİMELER: R programlama dili, apriori algoritması, veri madenciliği, birliktelik kuralı, Sepet analizi, yapay sinir ağları

ABSTRACT

MSc. Thesis

CUSTOMER BASED MARKET BASKET ANALYSIS WITH APRIORI ALGORITHM AND PRODUCT SALES FORECAST

Furkan OZTEMIZ

Inonu University

Graduate School of Natural and Applied Sciences

Department of Computer Engineering

88+x

2017

Supervisor : Prof. Dr. Serdar Ethem HAMAMCI

The popularisation and the improvement of data mining have enabled to attain significant patterns from ample data stacks. Numerous sectors have applied data mining methods on their data vigorously. In this thesis study, analysis of high quantity sales data belonging to a private firm providing services in the retail sector was made using apriori algorithm.We can state that the main purposes of analysis are to meet the customer satisfaction, to deepen commitment to the company, to raise the company's sales chart. By determining the association rules between the products bought by the customers, it was provided to identify the customer groups representing these rules. By this means, the company is provided to offer a few products to the right customers by organising numerous campaigns instead of generalising the customers by launching a single campaign. These special campaigns are aimed to get the customers feel special and to increase their commitment to the company. As a result of the analysis obtained from association rule methods, sales amount estimations of the products presented in the campaigns prepared for the customers were done through artificial neural networks. So substantial information were obtained about the extent of the customer interest to the campaign. In this analysis process, recording of the data systematically and bringing them in compliance with the analysis were carried out via MSSQL Server Database System. To establish the associatin rules, R programing language and R Studio setting were used. A priori algorithm was preferred as the method. In the estimation process carried out by applying YSA MATLAB was used. For determinig customers who represent the association rules dashboards were designed and the results are provided to have an easier and more comprehensible form.

KEYWORDS: R programing language, Apriori algorithm, Data mining, Association rules, Basket analysis, Artificial neural networks

TEŞEKKÜR

Yüksek Lisans tez çalışmamın her aşamasında yardım, öneri ve desteğini esirgemeden beni yönlendiren danışman hocam Sayın Prof. Dr. Serdar Ethem HAMAMCI'ya, bu çalışmanın analiz aşamalarında fikirleriyle katkı sağlayan Sayın Prof. Dr. Ali KARCI' ya ve Doç. Dr. Davut HANBAY' a, ayrıca tüm hayatım boyunca olduğu gibi yüksek lisans çalışmalarım sürecinde de benden desteklerini esirgemeyen değerli AİLEM 'e

teşekkür ederim.

Bu tez çalışması, İnönü Üniversitesi Bilimsel Araştırma Projeleri Koordinasyon Birimince 2016/22 proje numarası ile desteklenmiştir.

İÇİNDEKİLER

ÖZET……..……...i
ABSTRACT…….……..……ii
TEŞEKKÜR…….…….iii
İÇİNDEKİLER…….…….……iv
ŞEKİLLER DİZİNİ ..…….…….……. vi
ÇİZELGELER DİZİNİ ..…… viii
SİMGELER VE KISALTMALAR LİSTESİ ……..…… x
1. GİRİŞ ……..…….…… 1
2. VERİ KAVRAMI ve VERİ MADENCİLİĞİ……..…….……..6
2.1. Veri Tabanı, Veri Ambarı ve OLAP ……..…….…….…….....6
2.2. Veri Madenciliği ……..……..……7
2.2.1. Veri madenciliğinde ön işleme teknikleri …...…….…...….12
2.2.1.1. Veri temizleme…….……..…..…..…13
2.2.1.2 Veri birleştirme ……....…….…….....13
2.2.1.3. Veri dönüştürme……14
2.2.1.4 Veri indirgeme ……..……14
2.2.2. Veri Madenciliği Modelleri …….……..……...14
2.2.2.1. Sınıflama ve regresyon…….…….15
2.2.2.2. Kümeleme ……..……..…..…..15
2.2.2.3. Birliktelik kuralları ve ardışık zamanlı örüntüler …….....…...17
2.3.3. Veri madenciliği araçları ….……..…….……..17
3. BİRLİKTELİK KURALI ……..……..……..20
3.1. Birliktelik Kuralının Matematiksel Kavramları……...…….21
3.2. Birliktelik Kuralı Algoritmaları …….……..23
3.2.1. AIS algoritması . . ……..……..23
3.2.2. SETM algoritması.……..……..24
3.2.3. Apriori algoritması ……..…….25
3.2.4. Apriori TID algoritması ……..…….……..……..30
3.3. Birliktelik Kuralı Kullanım Alanları ……..……. 31

4. MÜŞTERİ PROFİLİNE DAYALI MARKET SEPET ANALİZİ UYGULAMASI ……...32

4.1. Analizde Kullanılacak Veri Seti ……..…….32

4.2. Müşteri Kartları……..……...35

4.3. Veri Hazırlama Programı …….36

4.4. R ile Apriori Algoritmasının Uygulanması……..…….…37

4.5. Birliktelik Kurallarına Ait Müşteri Bilgilerinin Tespiti……43

4.6. Analiz Sonuçlarından Elde Edilen Bazı Çıkarımlar …….…....64

4.7. Farklı Şubelere Ait Birliktelik Kuralları……..…….67

5. MÜŞTERİ BAZLI KAMPANYA OLUŞTURMA ……..68

5.1. Belirlenen Müşteri Kitlesine Özgü Oluşturulan Birliktelik Kuralları ….……..…68

5.2. Belirlenen Birliktelik Kurallarına Özgü Müşteri Profili Tespiti……...…71

6. BİRLİKTELİK KURALLARI İLE TESPİT EDİLEN KAMPANYA……75
ÜRÜNLERİNİN YSA ‘İLE SATIŞ MİKTARLARININ TAHMİN EDİLMESİ

7. TARTIŞMA VE SONUÇ …….……....81

KAYNAKLAR ……..….. 84

ÖZGEÇMİŞ …… 88

ŞEKİLLER DİZİNİ

Şekil 2.1. Veri madenciliği kullanım alanları ……9

Şekil 3.1. Klasik apriori algoritması özet kodu……..26

Şekil 3.2. Apriori algoritması akış diyagramı…….……..…….27

Şekil 3.3. AprioriTid algoritması sözde kodu……31

Şekil 4.1 Ürün satış hareketleri……..……33

Şekil 4.2. Ürünlerin hiyerarşik alt grupları…….…34

Şekil 4.3 Müşteri bilgileri……...34

Şekil 4.4. Müşteri analizi veri hazırlama programı…….36

Şekil 4.5. Frekansı en yüksek ürünler…….38

Şekil 4.6. Transaction'daki farklı Kod_3 grup sayıları……..38

Şekil 4.7. Ortalama Kod_3 grup sayıları……38

Şekil 4.8. En yüksek sayıda bulunan 10 ürün…….…38

Şekil 4.9. Destek, güven ve lift istatistikleri……39

Şekil 4.10. Müşteri yaş dağılım karşılaştırması……..…45

Şekil 4.11. Belirtilen ürün gruplarına ait yaş aralığı(Tüm Şubeler)……46

Şekil 4.12. Farklı marketlerdeki eğitim durumu karşılaştırması…….48

Şekil 4.13. Tüm şubeler eğitim seviyesi grafiği…….49

Şekil 4.14. Farklı şubeler meslek oranları karşılaştırma……51

Şekil 4.15. Tüm şubelere ait meslek grupları grafiği……..52

Şekil 4.16. Farklı şubeler müşteri cinsiyet oranları karşılaştırma……53

Şekil 4.17. Farklı şubelerdeki müşterilerin medeni durumları……53

Şekil 4.18. Farklı şubelerde belirtilen ürün grubunun tercih edildiği ayların ……..55
karşılaştırılması

Şekil 4.19. Farklı şubelerde belirtilen ürün grubunun tercih edildiği haftanın günlerine... 56
göre karşılaştırılması

Şekil 4.20. Farklı şubelerde belirtilen ürün grubunun tercih edildiği saat aralıklarına…...57
göre karşılaştırılması

Şekil 4.21. Farklı şubeler sepet tutar oranları karşılaştırma ……..59

Şekil 4.22. Belirlenen ürünlerin sepet içerisindeki oranlarının karşılaştırılması……61

Şekil 4.23. Şube 1003 adres oranları ……..63

Şekil 4.24. Şube 1004 adres oranları……..63

Şekil 4.25. Farklı şubelerdeki destek ve güven değerleri……66
Şekil 5.1. 1.Yıl belirtilen ürünler için ramazan ayındaki müşteri bilgileri……..….73
Şekil 5.2. 2.Yıl yılı belirtilen ürünler için ramazan ayındaki müşteri bilgileri…….74
Şekil 6.1. Genel YSA yapısı ……..….76
Şekil 6.2. Tasarlanan yapay sinir ağı ……..78
Şekil 6.3. Regresyon eğrileri ……..79

ÇİZELGELER DİZİNİ

Çizelge 3.1. Örnek müşteri alışverişleri…….22

Çizelge 3.2. Müşteri alışverişleri……26

Çizelge 3.3. Tekil ürünlerin tercih edilme sayısı……28

Çizelge 3.4. İkili ürün grubu destek sayısı……..28

Çizelge 3.5.Eşik değerinin üstünde kalan ürün grupları……..29

Çizelge 3.6. Üçlü ürün grupları destek sayıları……...29

Çizelge 3.7 Birliktelik kuralları……..30

Çizelge 3.8. 2000-2014 yılları arasındaki birliktelik kurallarına yönelik yapılan….31 makale sayısı

Çizelge 4.1. Veri istatistikleri……32

Çizelge 4.2. Birbirinden farklı ürün ve ürün grubu sayıları……33

Çizelge 4.3. Şube 1003 birliktelik kuralları……40

Çizelge 4.4. Şube 1003'de bisküvi ürün grubuna ait birliktelik kuralları……41

Çizelge 4.5. Şube 1004'e ait birliktelik kuralları……42

Çizelge 4.6. Şube 1004'de bisküvi ürün grubuna ait birliktelik kuralları……43

Çizelge 4.7. Şube 1003 ve 1004 'e ait en yüksek destek değere sahip birliktelik kuralları..44

Çizelge 4.8. Şube 1003 müşteri eğitim oranı……..47

Çizelge 4.9. Şube 1004 müşteri eğitim oranı……..47

Çizelge 4.10. Bütün şubeler eğitim oranları……47

Çizelge 4.11. Şube 1003 meslek grupları……....50

Çizelge 4.12. Şube 1004 meslek grupları……..……..50

Çizelge 4.13. Bütün şubeler meslek grupları……..…….50

Çizelge 4.14. Şube 1003 saat aralıklarına göre tercih oranları…….54

Çizelge 4.15. Şube 1004 saat aralıklarına göre tercih oranları…….54

Çizelge 4.16. Şube 1003 sepet tutar oranları……...58

Çizelge 4.17. Şube 1004 sepet tutar oranları…….….58

Çizelge 4.18. Tüm şubeler sepet tutar oranları…….…..58

Çizelge 4.19. Şube 1003 belirlenen ürünlerin sepet içerisindeki oranı……..60

Çizelge 4.20. Şube 1004 belirlenen ürünlerin sepet içerisindeki oranı……..60

Çizelge 4.21. Tüm şubeler belirlenen ürünlerin sepet içerisindeki oranı……60

Çizelge 4.22. Şube 1003 müşteri adres bilgileri…….62

Çizelge 4.23. Şube 1004 müşteri adres bilgileri……..62

Çizelge 4.24. Şubelerdeki belirtilen ürün gruplarının sepetteki ortalama tutarları……..64

Çizelge 4.25. Şubelere ait birliktelik kuralları……67

Çizelge 5.1. 1. yıla ait doktor olan müşteriler için birliktelik kuralları……..68

Çizelge 5.2. 2. yıla ait doktor müşteriler için birliktelik kuralları……..69

Çizelge 5.3. 1. yıla ait öğretmen olan müşteriler için birliktelik kuralları……..69

Çizelge 5.4. 2. yıla ait öğretmen olan müşteriler için birliktelik kuralları……..69

Çizelge 5.5. 1. yıla ait ev hanımı olan müşteriler için birliktelik kuralları…….……69

Çizelge 5.6. 2. yıla ait ev hanımı olan müşteriler için birliktelik kuralları…….……69

Çizelge 5.7. 1. yıla ait öğrenci olan müşteriler için birliktelik kuralları…….……69

Çizelge 5.8. 2. yıla ait öğrenci olan müşterileri için birliktelik kuralları……70

Çizelge 5.9. 1. yıla ait yüksek lisans mezunu olan müşteriler için birliktelik kuralları……70

Çizelge 5.10. 2. yıla ait yüksek lisans mezunu olan müşteriler için birliktelik kuralları…70

Çizelge 5.11. 1. yıla ait itibari ile belirlenen tarih aralığında alışveriş yapan müşterilere ait birliktelik kuralları……..……70

Çizelge 5.12. 2. yıla ait itibari ile belirtilen tarih aralığında alışveriş yapan müşterilere ait birliktelik kuralları……..70

Çizelge 5.13. 1. yıla ait Ramazan ayı birliktelik kuralları……71

Çizelge 5.14. 2. yıla ait Ramazan ayı birliktelik kuralları……71

Çizelge 5.15. 1. yıla ait belirtilen ürün gruplarındaki ürünlere ait birliktelik kurallar…….72

Çizelge 5.16. 2. yıla ait belirtilen ürün gruplarındaki ürünlere ait birliktelik kuralları……72

Çizelge 6.1. Birliktelik kuralları ……77

Çizelge 6.2. Ürün kümesi satış verileri ……..……77

Çizelge 6.3. Ürün kümesi satış verileri normalize edilmiş hali ……..78

Çizelge 6.4. Tahmin sonuçları ……...…79

SİMGELER VE KISALTMALAR LİSTESİ

$A \Rightarrow B$	: A ürünün bulunduğu satışlarda B ürünün de bulunması olayı
$A \subseteq T$	: A, T'nin alt kümesi
$A \cap B$	: A kesişim B
$A \cup B$	: A birleşim B
Ck	: K adetli sık geçen aday veri setleri
Lk	: Sık geçen k adet öğeli veri setleri
$Lk \infty Lk$	: K öğeli veri setlerinin kombinasyonları
MSSQL	: Microsoft SQL Server
YSA	: Yapay Sinir Ağları
YSH	: Yapay Sinir Hücresi

1.GİRİŞ

Son zamanlarda hemen hemen bütün sektörlerde üretimden, satış aşamasına kadar neredeyse bütün süreçlerde donanımsal ve yazılımsal otomasyon sistemleri kullanılmaktadır. Bu yapının bir gereği olarak ortaya çıkan veri boyutu'da yüksek oranda artmıştır. Bu denli yüksek boyutlu verilerin oluşması önemli bir problemi karşımıza çıkarmaktadır [1]. Bu veriler sade halleri ile pek işe yaramayan ve değerlendirilemeyecek kadar karışık bir şekilde sistemimizde gereksiz olarak yer tutmaktadır. Bu yorumlanamayan verilerin daha yönetilebilir, anlaşılabilir ve faydalı hale dönüştürülmesi için çeşitli analiz teknikleri geliştirilmiştir. Bu analiz teknikleri ve yöntemleri veri madenciliği çatısı altında toplanmıştır.

Veri madenciliği, veri işleme sonrasında ortaya çıkan sonuçların daha anlaşılır ve yorumlanabilir olması, ayrıca ilgili verilerden elde edilen tahminsel sonuçların gerçek sonuçlara hatırı sayılır derecede yakınlık sağlaması nedeniyle günümüzde yaygın olarak kullanılmaktadır. Veri madenciliği yöntemine duyulan ihtiyaçlar farklı benzetim tekniklerinin geliştirilmesinin önünü açmıştır. Veri madenciliği yöntemleri birçok sektörde hem raporlama hem de benzetim işlemleri ile beraber aktif olarak kullanılarak verilerin daha anlaşılabilir hale getirilmesinde ve daha faydalı verilere dönüştürülmesinde kullanılır [1]. Perakende ve toptan ticaret, sağlık, ulaştırma, taşımacılık ve bankacılık gibi birçok sektörde veri madenciliği yöntemleri sıkça kullanılmaktadır [2].

Veri madenciliği yöntemleri 4 ana alt başlık altında ifade edilebilir [3,4]:

1.Sınıflandırma

2.Kümeleme

3.Birliktelik Kuralları

4.Ardışık Zamanlı Örüntüler

Sınıflandırma işlemi veri madenciliğinde oldukça yoğun kullanılan bir yöntemdir ve verilere ait gizli örüntülerin ortaya çıkarılmasında tercih edilir. Öncelikle var olan verilerin bir kısmı sınıflandırma kurallarının oluşturulmasında kullanılır. Sonrasında ise bu kurallar doğrultusunda yeni oluşan bir durum için nasıl karar verileceği belirlenir. Kümeleme işlemi verilerin birbirleri arasındaki benzerlikleri tespit ederek bu doğrultuda gruplandırılmasıdır.

Birliktelik Kuralı verilerin birbirleriyle olan ilişkilerini inceler ve hangi verilerin birlikte olarak bulunabileceğine dair oransal bilgiler üreten veri madenciliği yöntemidir.

Ardışık zamanlı örüntüler meydana gelen bir dizi olay içinde en sıklıkla rastlanan alt dizilerdir.

Market sepet analizi müşterilerin alışveriş eğilimlerinin belirlenmesi ve bu eğilimlerin yorumlanabilir bir düzeye getirilmesi için uygulanan bir analiz çalışmasıdır. Satın alınan ürünlerin müşteriler tarafından hangi sıklıkla tercih edildiğinin çıkarılması ve bu ürünlerin hangi ürünlerle beraber satın alındığı bilgilerine erişmek satıcı firma tarafından önemli veriler içerebilir. Müşteri memnuniyeti hedeflenerek yapılacak olan market sepet analizi çalışması aynı zamanda firma tarafında marketlerdeki ürün stoklarının ürün gruplarına dönüştürülerek müşterilerin ürünlere daha rahat erişimini sağlamak amacıyla reyonlardaki ürün dizilimlerinin yapılması ve müşterilerin yapacakları alışveriş tercihlerinin önceden tahmini ile onlara uygun kampanya ve fırsat ürünlerinin sunulması gibi satış miktarını artırıcı birçok avantaj getirecektir.

1.1 Literatür Taraması ve Değerlendirilmesi

Birliktelik kuralına ait birçok farklı algoritma geliştirilmiştir. Bu algoritmaların birbirlerine karşı bazı farklılıkları vardır. Bazısı hız açısından üstünlük sağlarken bazısı da kullandığı donanım ve kolay uygulanabilirlik açısından tercih edilmektedir. Farklı sektörlerde uygulanan birliktelik kuralı yöntemlerine ait bazı literatür çalışmaları aşağıda belirtilmiştir [5].

- İstanbul Üniversitesi İşletme Fakültesinde müşterilerin alışveriş eğilimleri ile alakalı davranış modellemesi çıkarmak için bir çalışma yapılmıştır [6]. Bu analiz çalışmasında IBM SPSS Climentine programı kullanılarak müşterilerin tercih ettiği ürün grupları analiz edilmiş ve sonraki alışverişleri tahmin edilmeye çalışılmıştır.
- Türkiye'de otomotiv sektöründe faaliyet gösteren bir yetkili servisin müşterilerine ait alış-veriş verileri göz önüne alınarak, Apriori ve FP-Growth algoritmaları ile ürünler analiz edilmiştir. Böylelikle müşterilerin satın aldıkları ürünler arasındaki birliktelik kuralları belirlenmek istenmiştir. Bu doğrultuda yetkili servisin karını artırmaya yönelik uygulanacak kampanya ve promosyonlara yön verilmeye çalışılmıştır [7].
- Afyon Kocatepe Üniversitesi İşletme Anabilim Dalında ürün kategorileri arasındaki satış ilişkisinin belirlenmesi için birliktelik kuralı uygulanarak bir doktora tezi yazılmıştır [8].
- Trakya Üniversitesi Fen Bilimleri Enstitüsünde bir yüksek lisans çalışmasında perakende sektöründe çalışan bir firmanın ürünler için birliktelik kuralı yöntemleri farklı veri kümeleri üzerine uygulanmıştır [9].

- Marmara üniversitesi fen bilimleri dergisinde Sigortacılık sektöründe Müşteri ilişkileri yönetimine yönelik bir çalışma yapılmıştır. Müşterilerin karakteristik özelliklerinin belirlenmesi üzerine yapılan analiz işlemlerinde müşterinin sadakatini ve memnuniyetini artırıcı hedefler belirlenmiştir. Müşterilerin hangi ürün gruplarını tercih ettiklerini ortaya çıkarmak için yapılan çalışmada Apriori algoritması kullanılmıştır. Elde edilen analiz sonuçlarından faydalanarak müşterilere yönelik daha etkin sonuçlar verecek satış kampanyaları ve pazarlama stratejileri geliştirilmesinin önü açılmıştır [10].
- Bu çalışmada veri madenciliği yöntemleri ile üstün yetenekli öğrencilerin ilgi alanlarını tahmin etmek ve bu öğrencilerin ilgi duydukları alanların belirlenmesi hedeflenmiştir. Üstün yetenekli öğrencilerin ilgi alanlarını tespit edebilmek için 10 sınıflandırma algoritması seçilmiştir ve sonuçlar değerlendirilerek en uygun algoritma seçilmiştir. Apriori algoritması kullanılarak Üstün yetenekli öğrencilerin ilgi duydukları alanlar belirlenmiştir ve bu sonuçlar doğrultusunda ders programlarının daha etkin düzenlenmesi gibi konularda fayda sağlamıştır [11].
- Kurumsal kaynak planlama yazılımına ait günlük hata loglarını toplayarak birtakım ön işleme işlemleri gerçekleştirip, elde edilen veri üzerinde birliktelik analizi çalıştıran bir yazılım geliştirilmiştir. Bu yazılım sayesinde, bir hatanın oluşmasına yol açan olaylar arasındaki ilişkiler birliktelik analizi algoritması ile tespit edilerek, bu sonuçları raporlayan bir sistem geliştirilmiştir. Hataya en çok sebep olan örüntüler tespit edilmiş ve görselleştirilmiştir [12].
- Sosyal ağların çok yaygın bir kullanım alanı vardır. Bu durum insanları olumlu ve olumsuz yönde etkilemektedir. Bu çalışmada, Fırat Üniversitesi Teknik Eğitim Fakültesi Bilgisayar Öğretmenliği Bölümü öğrencilerinin sosyal ağları kullanma seviyeleri incelenmiş ve sosyal ağların öğrencileri ne derecede etkilediği tespit edilmeye çalışılmıştır. Bu nedenle öğrencilere bir anket formu yöneltilmiş ve elde edilen verilere veri madenciliği tekniklerinden biri olan birliktelik kuralı uygulanarak bazı kurallar elde edilmiştir [13].
- Bir yayınevi firmasının internet sitesindeki veriler göz önüne alınarak, veri madenciliği ve birliktelik kuralları tekniği ile sepet ve sipariş tabloları incelenmiştir. Hangi ürünlerin kategorilerinin değiştirilmesi gerektiği, kullanıcıların meslek ve ilgi alanı dağılımları, müşteri ilgi alanlarına göre satış grafikleri ve kullanıcıların ödeme seçenekleri ile ilgili bir veri madenciliği uygulaması gerçekleştirilmiştir [14,5].
- İnternet üzerinden ulaşılabilen basın yayın kaynaklarında yer alan görsel ve metinsel verilerin hızlı ve etkin bir şekilde erişimi ve bu kaynaklardan anlamlı ve önemli bilgilerin çıkarılması hedeflenmiştir. Çalışmalar istihbarat açısından önem

taşıyan kişi ve örüntülerle ilgili haberler üzerinde yoğunlaşmıştır. Çalışma Birliktelik kuralı yöntemlerinden apriori algoritması uygulanarak yapılmıştır. Bu uygulamadaki amaç kullanıcıların arama yapmak istedikleri kişilerin isimlerini yazarken yapabilecekleri yazım hatalarını tespit edip bunları elemektir [15,5].

Bu tez çalışmasında küçük-orta ve büyük ölçekte birçok şubesi olan büyük bir süper marketler zincirine ait satış ve müşteri verileri kullanılarak market sepet analizinin yapılması hedeflenmiştir. Bu veriler birçok şubesi bulunan süpermarketler zincirine ait 2 senelik satış ve müşteri verileridir. Analizi yapılacak veriler filtrelenerek sadece müşteri kartları ile yapılan alışverişlerden göz önüne alınarak seçilecektir. Analizde kullanılacak veriler bütün bir yılı kapsayacak şekilde seçilmiştir, çünkü müşteriler tarafından tercih edilen ürünlerin bütün mevsimlerdeki durumlarının çıkarılması hedeflenmiştir. Müşterilerin alışveriş verilerinin müşterilerin kişisel bilgileri ile ilişkilendirilmesi firma tarafından yapılacak herhangi bir promosyon ürün veya farklı kampanyaların hedef kitlesinin belirlenmesinde önemli avantajlar sağlayacaktır. Müşterilerin satın aldıkları ürünlerin takibi müşteri kartları ile yapılacaktır. Müşteri kartları müşteriye ait kişisel verilerin kayıt altına alınmasında kullanılacağı gibi aynı zamanda da müşterilere özel yapılacak kampanyalarda kullanımı mümkün olmaktadır. Market sepet analizi sonuçlarının müşterilere ait kişisel veriler ile birleştirilmesi ile satın alınan ürün gruplarının hangi yaşlardaki, hangi meslek grubundaki, hangi semtlerde yaşadıkları ve hangi öğrenim durumuna sahip insanlar tarafından yapıldığını anlamamızı sağlayacak veriler elde edilecektir. Elde edilecek bu çıkarımlar yapılacak kampanyaların hedef kitlesinin daha net belirlenmesini sağlayacaktır. Ayrıca birliktelik kuralı yöntemi kullanılarak tespit edilen birbirleri ile bağlantılı ürünlere ait yapılacak kampanyaların başarısının önceden tahmin edilmesine yönelik bir çalışma yapılmıştır. Yapay sinir ağları teknikleri kullanılarak gerçekleştirilecek bu çalışma sayesinde satışa sunulacak kampanyalı ürünlerin hangi şubede müşteriler tarafından hangi miktarda tercih edileceğine dair yapılacak başarılı tahmin sonuçları ile hem kampanyadan elde edilecek kârın önceden belirlenmesi hem de ilgili ürünlerin arz ve talep dengelerinin sağlanması için önemli sonuçlar elde edilecektir. Birliktelik kuralı yöntemlerinden apriori algoritması bu çalışma için kullanılacaktır. Analiz işlemi R Studio IDE ortamında ve R programlama dili ile yapılacaktır. Yapay sinir ağları ile yapılacak satış tahminleri için ise MATLAB programının NNtool modülü kullanılacaktır. Sonraki aşamada ise analiz sonuçlarının kümelenmesi ve yorumlanması için rapor işleminin gerçekleştirileceği grafiksel bir uygulama hazırlanacaktır. Bu uygulamadan müşterilere ait veriler başta olmak üzere, satış işlemine ait ayrıntılı analiz raporları alınabilecektir.

Bu tez çalışmasında Apriori algoritması kullanılarak özel bir firmadan alınan ürün satış verileri analiz edilerek ürün grupları arasındaki birliktelik kuralları tespit edilmiştir. Ayrıca birliktelik kurallarından faydalanarak oluşturulan ürün gruplarının satış miktarlarının tahmin işlemlerinde ise yapay sinir ağı teknikleri uygulanmıştır. Birliktelik kuralı üzerine literatürde çeşitli çalışmalar rapor edilmiştir. Ayrıca farklı sektörlerde farklı çalışma alanlarında birliktelik kuralı algoritmaları yoğun olarak kullanılmaktadır [5]. Bununla birlikte bu yöntem toptan ve perakende sektöründe kullanım sıklığı nedeniyle "market sepet analizi" veya "müşteri sepet analizi" olarak bilinmektedir.

2.VERİ KAVRAMI ve VERİ MADENCİLİĞİ

Herhangi bir işlemden geçmemiş ve tek başına yorum yapılamayacak kadar sistemsiz bir halde bulunan ham bilgi yığınına veri denir[16]. Ölçüm, sayım, deney, gözlem ya da araştırma yöntemleri ile elde edilen veriler nicel ve nitel olmak üzere 2 başlık altında gruplandırılabilir. Nicel veri sayılabilen, ölçülebilen ve sayısal bir değer bildirirken nitel veri sayısal bir değer bildirmeyen niteliklerin belirtildiği verilerdir. Verinin belirli bir hedefe yönelik olarak işlenmesi ile faydalı bilgi elde edilir. Günümüz teknolojisinin gelişmesi sonucu elde edilen büyük veri yığınları içinde faydalı bilginin elde edilmesi önemli bir problem olarak önümüze çıkmaktadır [17].

2.1. Veri Tabanı, Veri Ambarı ve OLAP

Veri tabanları birbiriyle bağlantılı bilgilerin depolandığı alanlardır. Günümüzde üretilen veri miktarları arttığı için depolanması gereken verilerin miktarı da doğru orantılı olarak artmıştır. Veri tabanı sistemleri büyük miktardaki verileri klasik yöntemlere alternatif olarak geliştirilmiştir. Son zamanlardaki veri tabanı yönetim sistemleri verileri sistemli bir şekilde kaydetme, silme, sorgulama ve taşıma imkânı sağlamıştır. Verilerin bu denli sistemli olarak kayıt altına alınması ile veri tabanı içerisinden istenen her türlü bilgiye erişim oldukça kısa sürede sağlanmaktadır. Güncel veri tabanı yönetim sistemleri aynı zamanda güvenlik, performans yönetimi, denetleme ve optimizasyon işlemlerinde önemli avantajlar sağlamaktadır [18]. Bazı önemli veri tabanı yönetim sistemleri şu şekilde sıralanabilir:

- MS-SQL Server
- Oracle
- IBM DB2
- PostgreSQL
- MySQL

MySQL ve PostgreSQL açık kaynak kodlu olup, ticari ve kişisel kullanım için herhangi bir lisans bedeli istememektedir. MS-SQL(MİCROSOFT), Oracle(ORACLE), DB2(IBM) veri tabanı yönetim sistemleri ise ticari kullanım için lisans bedeli istemektedir. Elde edilen veri yığınlarından faydalı bilgi elde edilebilmesi için verilerin doğru depolanması, doğru sınıflandırılması, doğru ayıklanması, doğru işlenmesi ve doğru yorumlanması gerekir. Son zamanlarda elde edilen veri miktarındaki artış nedeniyle, veri depolarının boyutları manuel olarak kontrol edilemeyecek düzeylere ulaşmıştır [8]. Veri ambarı birbiri ile bağlantılı verilerin sorgulanabildiği ve büyük analizlerin yapılabildiği bir veri deposudur. Veri ambarları sistematik olarak büyük miktardaki verilere kolay ulaşım imkânını sağlayacak bir yapıdır. Bu kavramı 1991 yılında öneren W.H. Immon'a göre veri ambarı verinin

temizlendiği, birleştirildiği ve yeniden düzenlendiği merkezi ve entegre bir depodur [9]. Kısaca verilerin yüksek boyutlarından dolayı normal veritabanı metotları ile çözümlenemeyecek analiz çalışmaları için geliştirilmiştir. Çok büyük miktarda veriler içerdiği için kullanıcılar için detaylı analiz sonuçları sağlayabilmektedir. Standart veri tabanı işlemlerinden daha hızlı sonuçlar vermektedir. Çünkü veri ambarında yer alan veriler sadece okunabilir bir yapıda bulunmaktadır. Güncellenme ve silinme gibi günlük işlemlere tabi tutulmaz, sadece yüksek raporlama işlemleri için tercih edilirler [3].

OLAP (On-Line Analytical Proccessing) stratejik konularda karar verme işlemini kolaylaştıracak bir şekilde veri ambarları üzerinde yapılan analiz ve sorgulama işlemidir. OLAP yapısı tüm veri tabanı sistemleri üzerinde yapılabilir. OLAP sorgulamaları normal veri tabanı sorgulamalarından farklıdır. Bir örnekle açıklamak gerekirse işlemsel bir veritabanında giriş ve çıkış işlemleri yani kayıt ekleme ve kayıt silme işlemleri yapılmaktadır. Bu verilerin market ürünleri olduğu varsayılırsa, normal sorgulama yöntemleri ile ürünlerin gün-ay-yıl gibi dönemlerdeki satış tutarları çıkarılabilir. Ürünlerin satın alma tutar ve miktarları, marketler arası transfer miktarları gibi değerler elde edilebilmektedir. Buna karşın OLAP sisteminde, sistemde kayıtlı olan veriler değerlendirilerek gelecekteki olası durumların ortaya çıkarılması mümkün kılınmaktadır. Örneğin, marketten süt ve peynir alan bir müşterinin tekrar ne zaman peynir alacağı, zeytin ve ekmek alan bir müşterinin bu ürün tercihleri içerinde domates satın alma ihtimali, veya belirli bir ürün grubunun belirlenen satış hedefini tutturma ihtimali OLAP yapısı ve sorgulamaları ile kolayca bulunabilir [19]. OLAP sistemleri normal raporlama sistemlerine göre oldukça hızlı çalışmaktadır. Bunun en önemli sebebi verilerin daha önceden hesaplanmış bir şekilde bulunmasıdır. OLAP verileri istenilen formattaki küp (cube) adı verilen ortamlarda tutulur [20]. Veri ambarındaki verilere OLAP sistemleri kullanılarak gerçekleştirilen çeşitli analiz işlemleri neticesinde Veri Madenciliği yapısı ortaya çıkmaktadır.

2.2 Veri Madenciliği

Artan teknoloji ile hemen hemen her sektörde büyük veri yığınları elde edilmektedir. Bu veriler tek başlarına ilk bakışta anlamsız ve yorumlanamaz olarak görülmektedirler. Çünkü bu kadar çok veri içerisinden ihtiyaç olan verilerin filtrelenmesi oldukça zor bir iştir. Normal istatistik yöntemleri ile ortaya çıkarılamayacak veri analiz işlemleri için veri madenciliği yöntemleri geliştirilmiştir. Veri madenciliği karmaşık ve düzensiz veriler içerisinden karar destek tabanlı değerli veriler üretme sürecidir [14,8].

Veri madenciliği tanımı literatür'de farklı şekillerde ifade edilmiştir:

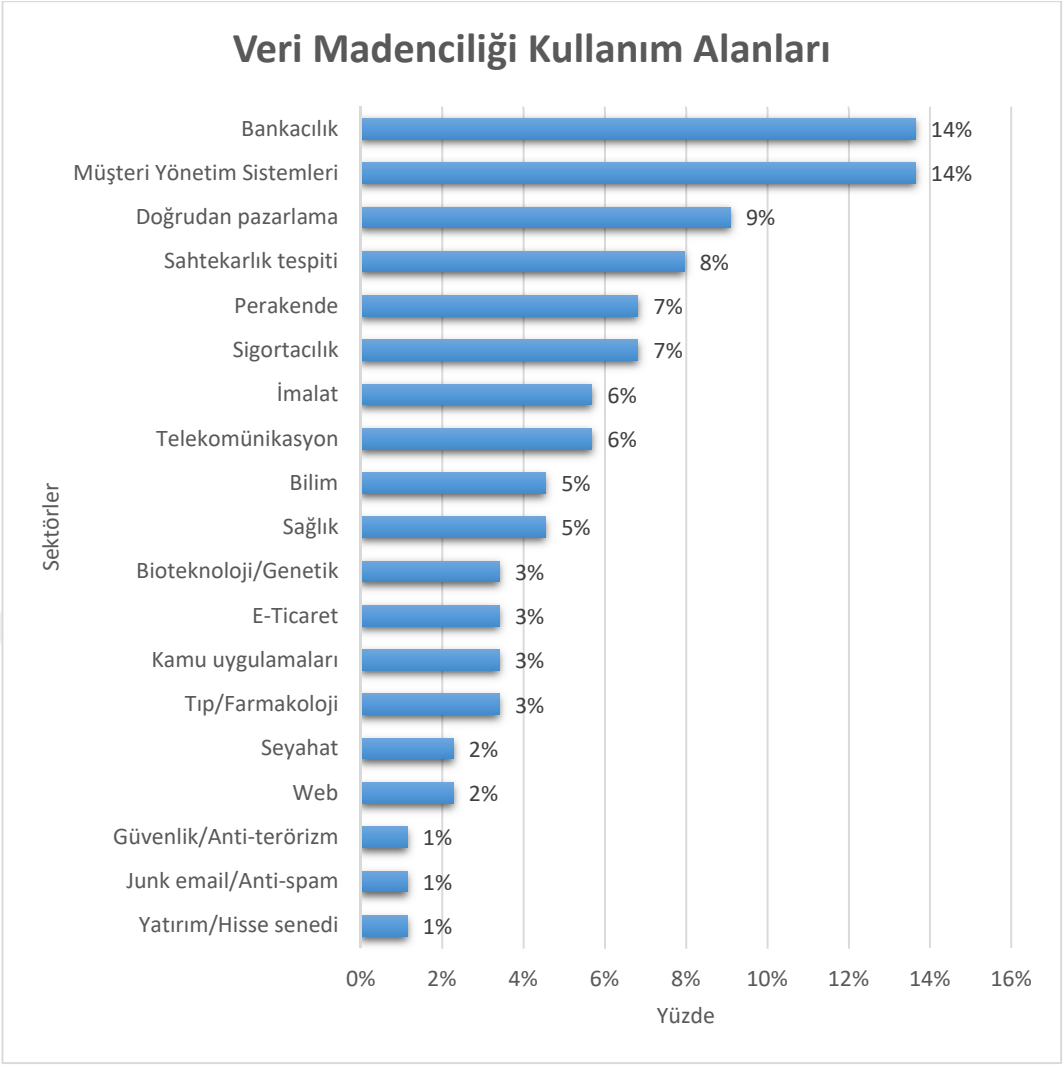
- Veri Madenciliği; veri içerisindeki trendleri, bağlantıları ve profilleri belirlemek için veriyi sınıflandıran analitik bir araçtır. Aynı zamanda ham verinin tek başına sunamadığı veriyi ortaya çıkaran veri analiz sürecidir [21].
- Veri madenciliği; mevcut veriler içerisinde net olmayan ve öncesinde bilinmeyen fakat potansiyel olarak kullanışlı bilginin elde edilmesi işlemidir [22].
- Veri Madenciliği; İşlenmemiş ham verinin tek başına istenen veya anlamlı şekilde gösteremediği bilgiyi ortaya çıkarabilecek veri analiz işlemleri olarak tanımlanmıştır [23].
- Veri Madenciliği; Veriyi üreten veya yorumlayacak kişiye veriyi daha anlamlı ve faydalı olacak şekilde, bu büyük veri yığınının daha önce var olduğu tespit edilmemiş ilişki bağlantılarının çıkarılması ve verinin metinsel bir şekilde özetlenmesi amacıyla yapılan analiz işlemidir [24].

Veri madenciliğinin temel özellikleri şu şekilde özetlenebilir [5]

- Büyük ve karmaşık veriler üzerinde uygulanırlar.
- Her tür sayısal veriler üzerine uygulanabilirler.
- Makine öğrenmesi, yapay zekâ, istatistik, veri tabanlarında bilginin keşfi, bilgisayar bilimleri, veri görselliği vb. gibi disiplinleri bulunmaktadır.
- Daha önceden tespit edilmemiş, doğrulanabilir bilgileri ortaya çıkarmayı amaçlar.
- Birçok sektörde kullanılmaktadır.
- Farklı sorulara göre değişen farklı çözüm araçları bulunmaktadır.
- Oldukça popüler ve hızlı büyüyen bir sektördür.

Veri madenciliği ile ilgili ayrıntılı bilgi için [3,19,25] nolu referanslar incelenebilir.

Veri madenciliğinin temel hedefi, çeşitli kaynaklar kullanılarak elde edilen verilerin anlamlı bilgilere dönüştürülmesi ve faydalı eylemlere dönüştürecek kararlar için kullanıma hazır hale getirilmesidir. Günümüzde veri madenciliğinin kullanım alanları gittikçe artmıştır. Kolay uygulanabilirliği ve ortaya çıkardığı etkili sonuçlar sayesinde birçok sektörde yaygın olarak kullanılmaktadır. Bu sektörlerin başında eğitim, ticaret, mühendislik, bankacılık, telekomünikasyon, perakende vb.. sektörler gelmektedir [5]. Şekil 2.1'de veri madenciliğinin çeşitli sektörler içinde kullanım oranları gösterilmiştir [9].



Şekil 2.1. Veri madenciliği kullanım alanları [9]

Tez çalışmasının konusunu da içeren birliktelik kuralları, yapay sinir ağları, vb. birçok veri madenciliği tekniğinin kullanıldığı bazı uygulama alanları şu şekilde listelenebilir [26];

- Ürün veya hizmette hangi özelliklerin ne derece müşterileri memnun ettiği, hangi özelliklerden dolayı bu hizmet veya ürünü tercih ettiği tespit edilebilir.
- Müşterilere ait kredi riskleri hesaplanıp, hangi müşterilerin kredinin geri ödemesini zamanın yapamayabileceği tahmin edilebilir.
- Kredi kartı ödemelerini zamanında gerçekleştirmeyen veya hiç ödemeyen müşterilerin özelliklerini göz önüne alarak bundan sonraki bu durumdaki muhtemel kişilerin tespitinde kullanılabilir.
- Piyasada oluşabilecek değişikliklere mevcut müşteri portföyünün vereceği tepkinin firma üzerinde yaratabileceği etkinin tespitinde kullanılabilir.
- Potansiyel müşteriler arasından en kârlı olanlar belirlenebilir, belirlenen müşteri gruplarına özel kampanyalar uygulanabilir.

- Bir ürün veya hizmetle ilgili kampanya oluşturulacağı zaman hedef kitlenin belirlenmesi aşamasından bu hedef kitleye kampanyaların nasıl sunulacağı kararına kadar olan bütün süreçlerde kullanılabilir.
- Geçmiş ve mevcut durum analiz edilerek geleceğe yönelik tahminler yapılabilir. Ciro tahminleri karlılık ve Pazar payı gibi analiz işlemlerinde kullanılabilir.
- Perakende sektöründe müşterilerin alışveriş tercihlerinin belirlenmesinde kullanılabilir.

Pazarlama sektörü ;

- Müşteri segmentasyonunda
- Müşterilerin demografik özellikleri arasındaki bağlantıların tespit edilmesinde
- Kampanya oluşturma ve yönetme işlemlerinde
- Sadık müşteri oluşturma çalışmaları
- Satış analizleri
- Müşteri ilişkileri yönetimi ve müşteri analizinde
- Satış tahminlerinde

Bankacılık sektörü ;

- Finansal göstergeler arasındaki gizli korelasyonların bulunmasında
- Kredi kartı dolandırıcılıklarının tespitinde
- Kredi taleplerinin değerlendirilmesinde
- Risk analizi ve risk yönetimi
- Sigortacılık, sigorta dolandırıcılık tespitinde
- Riskli müşterilerin tahmin edilmesinde

Perakendecilik sektörü ;

- Satış analizi
- Alış-veriş sepet analizi
- Müşteri kartı analizi
- Mağaza ürün yerleşim düzenlemesi

Borsa alanı ;

- Hisse senedi fiyat tahminleri
- Piyasa analiz işlemlerinde
- Hisse senedi alım-satım stratejilerinin belirlenmesinde

Telekomünikasyon sektörü ;

- Kalite ve iyileştirme analizlerinde
- Hatların yoğunluk tahminlerinde
- İletişim yapılarının belirlenmesinde
- Kaynakların daha verimli kullanılmasında
- Hizmet kalitesinin arttırılmasında

Sağlık sektörü ;

- Test, tahlil vb.. sonuçların tahmini
- Ürün geliştirme
- Hastalık teşhis işlemlerinde
- Tedavi sürecinin belirlenmesinde

Endüstri alanı ;

- Kalite kontrol analizler işlemlerinde
- Lojistik
- Üretim süreçlerinin belirlenmesinde

Literatürde veri madenciliği üzerine yapılmış çok sayıda çalışma bulunmaktadır. Bunlardan bazıları şu şekilde özetlenebilir;

- 3 fazlı asenkron motordaki sargı spirleri arasında oluşabilecek kısa devre veya yalıtım bozuklukları ve mekanik hataların tespitini gerçekleştiren bir çalışma yapılmıştır [27].
- Veri madenciliği ile deprem verilerinin analizi üzerine bir çalışma gerçekleştirilmiştir. Seçilen bir bölgeye ait sismik verilerin incelenerek gerçekleşme ihtimalinin veri madenciliği yöntemleri ile ortaya çıkarılmasını amaçlamaktadır. Örneğin belirlenen bir bölge için 5 magnitüdündeki bir depremin olma olasılığı 10 yıl için %27 iken 30 yıl için %60 değerlerini bulmaktadır tarzında çıkarımlar elde edilmiştir [28].
- İnternet üzerinden ulaşılabilen basın-yayın kaynaklarında yer alan görsel ve metinsel verilerin hem hızlı hem de etkin bir şekilde erişimi ile bu kaynaklardan anlamlı ve önemli bilgiler çıkarılması hedeflenmiştir. Çalışmalardaki analizi yapılan haberler daha çok istihbarat açısından önem taşıyan kişi ve örgütlerle ilgili seçilmiştir [15].
- Bir yazılımın güncel sürümlerindeki hata oranlarını önceki sürümlerine kıyasla değişikliklerini temel alarak tahmin eden bir model ortaya koymayı amaçlayan bir uygulama gerçekleştirilmiştir. Yazılımdaki yenilik algoritma değişikliği veya hata

ayıklama değişikliği olabilir. Bu tarz değişiklikleri analiz ederek yeni versiyonun hata oranını doğru bir şekilde tahmin edebilme amaçlanmıştır [29].

- Apriori algoritmasının ürettiği kurallar elenerek bir elektronik firmasında üretim ve mal giriş kalite verileri üzerinde uygulanmıştır. Ortaya çıkarılan kurallar test verileri ile doğrulanmış ve sonuçlar analiz edilmiştir [30].
- Kan biyokimya parametreleri ile demir eksikliği anemisi teşhisinde, hekime yardımcı olacak ve kolaylık sağlayabilecek bir karar destek sistemi oluşturulmuştur. Karar destek sisteminin sonuçları ile Anemi teşhislerinde önemli neticeler elde edilmiştir [31].
- İMKB 100 endeksinde sanayi ve hizmet sektörlerinde faaliyet gösteren 173 işletmenin 2004–2006 yıllarına ait yıllık finansal göstergelerinden yararlanarak veri madenciliği tekniklerinden birisi olan karar ağaçları tekniği uygulanmıştır. Seçilen finansal göstergelere göre sanayi ve hizmet sektörlerinde faaliyet gösteren firmaları ayıran en önemli değişkenler saptanmıştır [32].
- İMKB'de işlem gören ve imalat sektöründe faaliyet gösteren 100 firmanın bilgilerini içeren kaldıraç oranı ve aktif karlılık oranının finansal tablo hilesini tespit etmek için bir çalışma yapılmıştır [33].
- ÖSYM tarafından 2008 ÖSS adayları için resmi internet sitesi üzerinden yapılan anket verileri üzerinde veri madenciliği yöntemleri kullanılarak, öğrencilerin başarılarını etkileyen faktörler araştırılmıştır. Bu araştırmada, veri madenciliği yöntemlerinden karar ağaçları ve kümeleme kullanılmıştır [34].
- Türkiye'de bir hava yolu işletmesinin parça söküm raporları üzerinde veri madenciliği çalışması gerçekleştirilmiştir. Çalışmanın amacı, uçaklarda kullanılan parçaların, herhangi bir arıza oluşmadan önce düzeltici ve önleyici işlemlerin yapılması için ikaz seviyelerinin tespit edilmesine yönelik kural geliştirmektir [35].

2.2.1 Veri madenciliğinde ön işleme teknikleri

Veri madenciliğinde en önemli konulardan bir tanesi analiz edilecek verinin kalitesidir. Analiz sonucunda elde edilecek verinin güvenilir ve doğru olması için verilerin ön işlemden geçirilmesi gerekmektedir. Eğer bu işlem yapılmazsa giriş verilerindeki hata oranı kadar elde edilecek sonuçta hatalı olacaktır. Özellikle son zamanlarda ortaya çıkan verilerin yüksek boyutlu olması veri madenciliği işlemlerinin uygulanmasında, veri ön işleme tekniklerinin kullanımını daha önemli bir hale getirmiştir [1]. Örneğin, bir firmanın müşterilerinin alışveriş eğilimlerini tespit edebilmek için müşterilere ait satış hareketlerini görmek istediği durum için, satış ve müşteri verilerinin ayrı veri tabanlarında olması ciddi performans ve zaman kayıplarına neden olabilir, ileride kullanılacağı düşünülmemiş olan bazı verilere sonradan ihtiyaç duyulabilir, verilerin bir kısmı silinerek veri tabanını tutarsız

bir yapıya getirmiş olabilir, zaman içerisinde yapılan kayıt-silme- güncelleme gibi işlemlerden dolayı veri tabanındaki ilk kurgulanan düzenli yapı bozulmuş olabilir. Vs. Bunun gibi birçok sorun çözülmeden uygulanacak olan veri madenciliği yöntemleri ile faydalı sonuçlar elde etmek mümkün değildir. Bu yüzden veri ön işleme tekniklerinin kullanımı ve süreci veri madenciliğinin önemli bir adımıdır [25].

Bir veri ön işleme tekniği dört önemli işlem adımı içermelidir [1,25,36]:

1. Veri temizleme işlemi
2. Veri birleştirme işlemi
3. Veri dönüştürme işlemi
4. Veri indirgeme işlemi

2.2.1.1 Veri temizleme

Veri tabanında hatalı veya tutarsız şekilde yer alan verilere gürültülü veri denilmektedir. Verilerdeki gürültüyü temizlemek için çeşitli işlemler uygulanabilir. Veri tabanında eksik değer içeren kayıtlar olabilir bunların çözülmesi gerekmektedir. Eksik olan değerler yerine sabit değerler atanabilir veya bu eksik değerler atılabilir. Diğer verilerin ortalaması alınarak eksik olan değerlerin yerine yazılabilir. Verilere (regresyon, karar ağacı, kümeleme, gruplama) vb. işlemler uygulayarak çıkan sonuç eksik olan veriler yerine kullanılabilmektedir.

2.2.1.2 Veri birleştirme

Birden fazla veri kaynağından alınan verileri birlikte değerlendirmek için verileri tek türe dönüştürmek veya tek kaynaktan alınması gerekebilir. Bu durumlarda veri birleştirme işlemini uygulamak gerekebilmektedir. Örnek ile açıklamak gerekirse; marketlerdeki satış işleminin gerçekleştirildiği kasalarda müşterilere ait veriler tutulmamaktadır. Müşterilere ait veriler arka tarafta farklı bir veri tabanında tutulabilmektedir. Kasada ürünleri satın alan müşteriye ait Müşteri_ID bilgisi veri tabanına gönderilmektedir. Veri tabanından ise Müşteri_ID'leri, KartID gibi farklı bir isimde tutulabilir. Bu noktada kasalarda müşterilere ait verilerin getirilmesi gerektiğinde(Fiş içeriğine müşterinin isminin yazılması gibi vb.) bu iki parametrenin birleştirilerek tek yapı halline getirilmesi gerekmektedir. Birleştirme işlemi uygulanırken iki veri kaynağı arasında tutarsız bulunan verilerin atılması işlemi de verinin doğruluğuna önemli ölçüde faydalar sağlamaktadır. Yeni elde edilecek bilgiler ve bu bilgilerin güvenilirliği mevcut verilerin birbiri ile uyumuna bağlı olmaktadır.

2.2.1.3 Veri dönüştürme

Verilerin veri madenciliğinde kullanılacak modele göre şeklinin dönüştürülmesi işlemidir. Yapılan dönüştürme işleminde verinin içeriği korunmalıdır. Veri dönüştürme süreci düzeltme, birleştirme, normalleştirme gibi farklı işlemler içerebilir. En popüler olan veri dönüştürme işlemi, verinin normalleştirilmesi işlemidir. Veri normalleştirme tekniklerinden bazıları Min-Max, Ondalık Ölçekleme, Z Skor vb. olarak sıralanabilir.

2.2.1.4 Veri indirgeme

Eğer veri madenciliği uygulamalarından sonra elde edilecek analiz sonuçlarının değişmeyeceği var sayılıyorsa, gerekli görülen veri veya değişkenlerin sayıları ve boyutları azaltılabilmektedir. Veri indirgeme yöntemleri aşağıdaki gibi belirtilmektedir ;

- Veri Birleştirme veya Veri Küpü
- Boyut indirgeme
- Veri Sıkıştırma
- Kesikli hale getirme

2.2.2 Veri Madenciliği Modelleri

Veri madenciliğinde farklı modeller kullanılmaktadır. Bu modeller tahmin edici ve tanımlayıcı olarak iki ana başlık altında toplanmıştır [26].

Tahmin edici modellerde, ilk aşamada sonuçları tespit edilmiş veriler baz alınarak bir model geliştirilmesi hedeflenmektedir. Daha sonrasında oluşturulan bu modelden faydalanılarak, sonuçları bilinmeyen veri kümeleri için sonuç değerlerinin tahmin edilmesi amaçlanmaktadır. Bir bankadan kredi alacak olan bir müşterinin bu krediyi ödeyebilip ödeyemeyeceğine dair verilerin tahmin edilmesi bu tür modele bir örnek olarak gösterilebilir. Yapılan tahmin neticesinde müşterinin geçmiş verileri incelenerek, müşterinin kredi notu belirlenmektedir ve müşterinin kredi karnesinin kredi alabilmek için uygun olup olmadığı tespit edilecektir [37].

Tanımlayıcı modellerde ise karar vermede yönlendirici olarak kullanılabilecek mevcut verilerdeki örüntülerin tanımlanması sağlanmaktadır. X/Y aralığında geliri ve iki veya daha fazla arabası olan çocuklu aileler ile çocuğu olmayan ve geliri X/Y aralığından düşük olan ailelerin satın alma örüntülerinin birbirlerine benzerlik gösterdiğinin belirlenmesi tanımlayıcı modellere bir örnektir [37].

Veri madenciliği yöntemlerinden sınıflama ve regresyon yöntemleri tahmin edici modeller olarak belirtilirken, kümeleme ve birliktelik kuralı yöntemleri ise tanımlayıcı modeller olarak karşımıza çıkmaktadır.

2.2.2.1 Sınıflandırma ve regresyon

Verilerin önceki çıktılara göre gruplandırılması işlemine sınıflandırma denir [19]. Sınıflama ve regresyon önemli veri sınıflarını belirleyen ve gelecek veri eğilimlerini tahmin eden modelleri kullanabilen iki veri analiz yöntemidir [37]. Geleceğin tahmin edilmesinde mevcut verilerden hareket ederek veri madenciliği modelleri arasında kullanılan en yaygın yöntemler sınıflama ve regresyon modelleridir. Sınıflama ve regresyon modelleri arasındaki temek fark tahmin edilen bağımlı değişkenin kategorik Sınıflama kategorik değerleri tahmin ederken, regresyon ise süreklilik gösteren değerlerin tahmin edilmesinde kullanılır [8].Sınıflama modelinde birçok uygulama örneği mevcuttur. Bunlardan bazıları;

- Kredi başvurusu değerlendirme(risklerin düşük-orta-yüksek olacak şekilde sınıflandırılması)
- Hastalık teşhisi
- Karakter tanıma
- Kullanıcı davranışları belirleme
- Hileli sigorta başvurularının belirlenmesi vb. gibi birçok örnek verilebilir [8].

Sınıflama ve regresyon modellerinde kullanılan başlıca teknikler aşağıdaki gibidir [37];

- Karar Ağaçları
- Yapay Sinir Ağları
- Genetik Algoritmalar
- K-En Yakın Komşu
- Bellek Temelli Nedenleme
- Naive-Bayes
- Regresyon Ağaçları

2.2.2.2 Kümeleme

Birbirleri ile benzerlik gösteren veriler içerisindeki bu benzer verilerin gruplandırılması işlemi olarak tanımlanabilir [38]. Her kayıt mevcut küme ile karşılaştırılarak ve kendisine en yakıt kümeye eklenmektedir. Eklenen yeni kayıt kümenin tanımlayıcı değerini değiştirerek, küme bir sonraki eklenecek olan kayıt için faklı küme merkezine sahip olmaktadır ve küme merkez değeri yeniden ayarlanmaktadır [21]. Kümeleme analizi büyük ölçekli sistemleri daha küçük bileşenlere ayırarak böl ve yönet metodolojisi üzerine kurulmuştur. Kümeleme analiz yöntemi heterojen yapıya sahip veri topluluğunu daha homojen alt gruplara ayrılmasını sağlayacaktır [8]. Sınıflama ve Kümeleme işlemleri birbirlerine benzer yapılar olarak görülmektedir. Çeşitli farklılıklara sahip bu iki analiz yöntemini birbirinden ayıran en önemli fark, Kümeleme analizinde

yapılacak işlem sınıflama işleminde olduğu gibi belirlenmiş yakın sınıfların durumlarına göre yapılmamasıdır. Sınıflama işleminde her veri daha önceden bir eğitim neticesinde oluşturulan modele göre bir sınıfa atanmaktadır. Kümeleme işleminde ise daha önceden tanımlanmış sınıflar bulunmamaktadır [8]. Kümeleme işleminde kümenin değerleri kendisine dahil olan verinin etkisiyle değişmektedir. Veriler birbirlerine olan benzerliklerine göre uygun kümelere eklenebilirler [8].

Kümeleme modelinde verilerin herhangi bir sınıfı bulunmamaktadır. Sınıflama modelinde veri sınıfları daha önceden belirlenmiştir ve yeni gelecek verilerin hangi sınıftan olacağı tahmin edilebilmektedir. Kümeleme modelinde ise verilere ait sınıflar bulunmamaktadır ve bu sınıflandırılmamış veriler gruplar halinde kümelere ayrılmaktadır [37]. Örneğin farklı semtlerde yaşayan müşterilerin market alışverişleri örüntülenebilir. Bu müşteri gruplarının hangi ürün gruplarını tercih ettikleri, bu ürünleri tercih ederken hangi saat aralığında markete gittikleri ve bunun yanında ürün sınıflandırmaları, market ve bulundukları semtlere ait sınıflandırmalar yapılması amacıyla kümeleme işlemi kullanılmaktadır.

Kümeleme analiz modelinin kullanım alanlarından bazıları aşağıdaki gibidir;

- Coğrafik sınıflandırma
- Perakende sektörü(Müşteri ve Ürünlere ait örüntülerin belirlenmesinde)
- Bilimsel sınıflandırma (Fen, Biyoloji ve Kimya gibi bilim dallarında yapılacak sınıflandırma işlemlerinde)
- Bilgi keşiflerinde (Bilgiye ulaşımı kolaylaştırmak için dokümanların sınıflandırılması)
- Tıp alanında (genetik araştırmalar gibi konularda sıklıkla kullanılır.)

gibi birçok alanda kümeleme analiz modeli kullanılmaktadır [8,37].

Literatürde birçok kümeleme modeline ait algoritmalar bulunmaktadır. Yapılacak analiz işleminde bu algoritmanın seçimi verinin tipine ve analizin amacına bağlı olarak değişmektedir. Kümeleme yöntemleri şu şekilde kategorilere ayrılmıştır [9];

- Bölümleme Yöntemleri
- Hiyerarşik Yöntemler
- Yoğunluk Tabanlı Yöntemler
- Izgara Tabanlı yöntemler
- Model Tabanlı yöntemler

2.2.2.3 Birliktelik kuralları ve ardışık zamanlı örüntüler

Bu veri madenciliği modelinde veri tabanındaki geçmiş verileri analiz edilerek geleceğe yönelik bilgiler elde edilmesi amaçlanmıştır [9]. Bir diğer deyişle birliktelik kuralı yöntemleri veri tabanı içerisindeki kayıtların birbirleri ile ilişkilerini belirler ve bu verilerden hangilerinin beraber olarak gerçekleştiğini ortaya çıkarmaktadır [9]. Birliktelik kuralları ve ardışık zaman örüntüleri arasındaki fark zaman kavramının uygulamaya dahil edilmesidir. İki yöntem de nesnelerin birbirleri arasındaki birliktelik bağlantılarını ortaya çıkarmayı amaçlamaktadır. Analiz işlemi sırasında belirli bir zaman aralığı temel alındığı zaman bu analiz işlemi "ardışık zamanlı örüntü çözümlemesi" şeklinde adlandırılmaktadır [39].

Daha çok perakende sektörü üzerinde uygulanan birliktelik kuralı yöntemleri ulaşım, sağlık, taşımacılık, bankacılık vb. gibi birçok sektörde kullanılmaktadır [40]. Birliktelik kuralı eş zamanlı olarak gerçekleşen ilişkilerin belirlenmesinde kullanılmaktadır [8]. Bazı örnek analizler sonuçları aşağıda belirtilmiştir;

- Beyaz peynir satın alan bir müşteri %80 ihtimal ile siyah zeytin de satın almıştır.
- Ceket ve pantolon satın alan müşteriler %50 ihtimal ile kravat da satın almıştır.
- Bir yere giderken A vasıtasını kullanan yolcular %15 ihtimal ile dönüş için yine A vasıtasını tercih edebilir tarzında değişik örnekler verilebilir.

Ardışık zamanlı örüntüler ise birliktelik kuralındaki gibi eş zamanlı olarak gerçekleşen değil birbirini izleyen belirli dönemlerde gerçekleşmiş olan ilişkilerin ortaya çıkarılmasında kullanılmaktadır [8]. Birkaç örnek analiz aşağıda belirtilmiştir;

- Katarak ameliyatı olan bir hastanın 10 gün içerisinde gözünün enfeksiyon kapma ihtimali %55 dir.
- Otomobil tamir servisinde motoru tamir edilen bir arabanın 30 gün içerisinde aynı arızayı verme ihtimali %10 dur.

vb.. gibi ardışık zamanlı örüntüler için birçok örnek verilebilir.

2.3.3 Veri madenciliği araçları

Veri madenciliği yöntemlerinin uygulanabilmesi için çeşitli yazılım çözümleri geliştirilmiştir. Bu geliştirilen programları birbirlerine göre farklı alanlarda farklı avantajlara sahiptir. Bu analiz programlarının ortak amacı; kararları ve sonuçları geliştirme, verilerden değer elde etmeye yardımcı olma, var olan sistem ile daha kolay bütünleştirme olarak belirtilebilir. Bunlardan bir kısmı Ticari, diğer bir kısmı ise açık kaynak kodlu yazılımlardır. IBM SPSS Modeler, MATLAB ticari yazılımlara örnek verilirken, RapidMiner, Weka, R, Keel yazılımları ise açık kaynak kodlu yazılımlara örnek verilebilir. Veri madenciliğinde kullanılan bazı araçlar;

❖ IBM SPSS Modeler

Integral Solutions tarafından Clementine ismi ile piyasaya sunulmuştur ve şuan IBM firması tarafından geliştirilmeye devam etmektedir. Kişiler, gruplar, sistemler ve kuruluş tarafından alınan kararlara tahmine dayalı zekâyı getirmek için tasarlanmış, tahmine dayalı bir analitik platformudur. Daha iyi sonuçlar sağlayan işlemleri seçmenize yardımcı olmak için metin analitiği, varlık analitiği, karar yönetimi ve optimizasyon da dahil olmak üzere bir dizi gelişmiş algoritma ve teknik sunar. Bulut tabanlı sürümü de dahil birkaç sürümü bulunan SPSS Modeler, masaüstünde devreye alımlardan işletim sistemleri içinde bütünleştirmeye kadar ölçeklenebilir [41].

❖ MATLAB

MATLAB (matrix laboratory), çok paradigmalı sayısal hesaplama yazılımı ve dördüncü nesil programlama dilidir. MathWorks tarafından geliştirilmektedir. Matlab kullanıcıya, matris işleme, fonksiyon ve veri çizme, algoritma uygulama, kullanıcı arayüzü oluşturma vb. birçok imkânı tanır. MATLAB, yaygın olarak akademik ve araştırma kurumlarında olduğu kadar endüstriyel işletmelerde de çok kullanılmaktadır [42].

❖ KEEL

Granada Üniversitesi tarafından geliştirilmiştir. Java dilinde geliştirilmiştir [43]. Klasik veri madenciliği algoritmaları açısından zengin değildir. Daha çok yapay zekâ tabanlı sınıflandırma işlemleri, kural tabanlı kümeleme algoritmalarının birçoğunu bünyesinde barındırmaktadır [F44]. Veri görselleştirme açısından zayıftır.

❖ RAPID Miner

Dortmund teknoloji üniversitesinde geliştirilmiştir. Veri analizi, Ön işleme, Sınıflama, Kümeleme, Birliktelik Kuralları Çıkarımı, Nitelik seçimi işlemlerini içerir. Oracle, MsSQL, PostgreSQL vb. birçok veri tabanını desteklemektedir [44]. Görselleştirme ve grafik arayüzü açısından en zengin yazılımlar arasındadır [45].

❖ WEKA

Waikato Üniversitesinde geliştirilmiştir. Alt yapısı java platformundan oluşmaktadır. İçerdiği özelliklerle veri kümeleri üzerinde önişleme, sını andırma, kümeleme, birliktelik kuralı madenciliği, özellik seçimi ve görselleştirme yapabilmektedir [46]. Java Database Connectivity ile SQL veri tabanlarına bağlantı sağlayabilir [47].

❖ R

Auckland Üniversitesi tarafından geliştirilmiştir. Grafikler, istatistiksel hesaplamalar ve veri analizleri için geliştirilmiş bir yazılımdır. Unix makinalarda yaygın

olarak kullanılmaktadır. R, kullanıcıların eklediği özel fonksiyonlar veya çok özel araştırma alanlarına ait paketlerle oldukça geliştirilebilirdir. Sahip olduğu paketlerin 3.parti yazılımlarla entegre olarak çalışması çok önemli bir avantaj olarak sunulmaktadır [48]. Bu tez çalışmasında R programlama dili ile birliktelik kuralı çalışması yapılacaktır.

- SAS Enterprise Miner

SAS Enterprise Miner, dağıtılmış istemci-sunucu sistemi olarak sunulmaktadır. Veri madencileri ve iş analistleri için açıklayıcı modeller oluşturmak ve iş süreçlerine dahil edilebilecek sonuçlar üretmek ve daha hızlı çalışabilmesi için optimize edilmiş bir mimari sağlamaktadır. Veri madenciliği sürecini geliştirmek için veri entegrasyonu, ön görme ve raporlama gibi diğer SAS teknolojileri ile uyumlu çalışacak şekilde tasarlanmıştır. R programlama dili entegre olabilmektedir ve gerçek zamanlı olarak web üzerinde, ilişkisel veri tabanlarında veya Hadoop üzerinde dağıtım imkanı sağlamaktadır [49][50].

3.BİRLİKTELİK KURALI

Birliktelik kuralları bir veri yığınındaki nesnelerin birlikte görülmelerinin olasılık yapılarıyla gösterilmesidir. Birliktelik kuralları verilerin potansiyel ilişkilerinin ortaya çıkarılmasıdır. Beraber gerçekleşen muhtemel olayların kurallarının belirlenmesini hedeflemektedir [24]. Birliktelik kuralı geçmişte elde edilen verilerin analiz edilerek bu veriler içindeki birliktelik durumlarının tespitini belirlemeye yönelik yapılan yöntemdir. 90'lı yıllardan bu yana hemen hemen birçok sektörde teknoloji aktif olarak kullanılmaktadır. Teknolojinin bu kadar yaygınlaşması ile ortaya çıkan verilerin kayıt altına alınma sürecini oluşturmuştur. Bu verilerin kayıt altına alınma süresi barkod, manyetik kart, Çip teknolojileri ile daha fazla hızlanmıştır. Bu tip gelişmeler günümüzde perakende sektörünü başta olmak alış işlemi ve satış işlemlerinin kayıt altına alınıp elektronik ortamlara aktarılmasına olanak sağlamıştır [9]. Elektronik ortamda yani veri tabanlarına kaydedilen bu veriler pazarlama alt yapısının önemli parçalarından birisidir. Bu verilerin depolanması ve işlenmesi için önemli çabalar harcanmaktadır [25]. Birliktelik kurallı yöntemi piyasada müşteri sepet analizi veya market sepet analizi olarak bilinen işlemlerde aktif olarak kullanılmaktadır [40]. Bizim çalışmamızda da birliktelik kuralı yöntemleri önemli bir yer tutmaktadır.

Birliktelik kuralında, müşterilerin satın aldıkları ürünler arasındaki birlikte bulunma durumlarının istatistik olarak ortaya çıkarılması ile müşterilerin satın alma eğilimlerinin belirlenmesini amaçlamaktadır. Yapılan analiz sonucunda üretilen birliktelik kuralları sayesinde satıcılar daha etkin ve kazançlı satışlar yapabilme imkânına sahip olunacaktır. Örneğin: Market alışverişleri sırasında zeytin aldıktan sonra yine aynı alışverişte %75 ihtimal ile peynir satın alıyorsa, bu iki ürün arasında birliktelik kuralı oluşturulabilmiştir. Elde edilen bu veri sayesinde bu ürünler üzerinde çeşitli kampanyalar yapılabilir. Yapılacak bu kampanyalar ile müşteri sayısında ve satışlarda ciddi artışlar sağlanabilir. Marketçilik sektöründe birliktelik kuralı yöntemleri kullanılarak elde edilen analiz sonuçları ışığında yapılacak kampanyalar ile elde edilecek bazı kazanımlar aşağıda listelenmiştir;

- Birliktelik kuralı tespit edilen ürünleri birlikte kampanyalı ürün olarak müşteriye sunmak
- Birlikte tercih edilen ürünlerin müşteriler tarafından kolay ulaşılabilir olması için market içerisindeki raf ve reyon dizilimlerini hazırlamak
- Firma'nın elindeki stokları hızlı eritme imkânı (Son kullanma tarihi yaklaşan ürünler için oluşturulan kampanyalar.)
- Müşterilerin alışveriş eğilimlerinin incelenmesi neticesinde müşteri sadakatinin belirlenmesi

- Müşterilerin tercihlerinin önceden belirlenmesi ile marketteki stok durumunun düzenlenmesi
- Müşterilerin kişisel bilgileri ile birliktelik kurallarının birleştirilmesi ile müşteriye özel indirim yapılabilecektir. Bu durum da müşterilerin sadakatini artırıcı etkiler sağlayacaktır.

3.1 Birliktelik Kuralının Matematiksel Kavramları

Birliktelik kuralları yöntemi ilk olarak 1993 senesinde Agrawal, Imelinskive ve Swami tarafından ele alınmıştır [51]. Birliktelik kuralları destek ve güvenirlik seviyelerine göre karakterize edilirler. Destek ve Güven değerlerinin yüksek olması bağıntıların güçlü olduğunu göstermektedir [52]. Kullanıcılar tarafından belirlenen eşik değerlerine göre genel birliktelik kuralları oluşturulabildiği gibi belli bir nesneye ait birliktelik kuralı bağlantıları da oluşturulabilir. Agrawal ve Srikant sınırlanmış nesne ile kurallar için çeşitli etkin çözümleme yöntemleri geliştirmişlerdir [53].

Birliktelik kurallarının temel bileşenleri Destek(Support) ve Güven(Confidence) değerleridir. Bu elde edilen değerler belirtilen eşik değeri ile kıyaslanarak istenen güçlü kuralların tespit edilmesini sağlar. Destek ve Güven değerlerine ait formüller aşağıda belirtilmiştir; A ve B farklı iki öğe kümesi olarak adlandırılabilir.

$$Destek(A) = \frac{A \text{ öğe } küme\ sayısı}{Toplam \text{ öğe } küme\ sayısı} \quad (3,1)$$

$$Destek(A,B) = \frac{(A,B) \text{ öğe } küme\ sayısı}{Toplam \text{ öğe } küme\ sayısı} \quad (3,2)$$

$$Güven(A,B) = \frac{(A,B) \text{ öğe } küme\ sayısı}{A \text{ öğe } küme\ sayısı} = \frac{Destek\ (A,B)}{Destek\ (A)} \quad (3,3)$$

(3,1),(3,2),(3,3) numaralı denklemler ile tanımlanmaktadır [54]. 3,1 numaralı(Destek(A)) denklem A öğesinin bütün öğeler kümesinde bulunma oranını bize vermektedir. 3,2 numaralı(Destek(A,B)) denklemi hem A hem de B öğelerinin birlikte bulundukları durumların bütün öğeler kümesine oranıdır. 3,3 numaralı denklem(Güven(A,B)) ise A öğesinin bulunduğu durumlarda B nesnesi ile birlikte bulunma ihtimalini vermektedir.

Çizelge 3.1. de müşterilerin yapmış oldukları örnek alışverişler verilmiştir.

Çizelge 3.1. Örnek müşteri alışverişleri

Alışveriş ID	Satın alınan ürünler
1	Ekmek, Peynir, Gofret, Yağ
2	Ekmek, Peynir, Gofret, Et
3	Ayran, Yağ, Peynir, Gofret
4	Peynir, Gofret, Yağ
5	Ekmek, Peynir, Gofret, Yağ

Yukarıda belirtilen alışveriş hareketlerine birliktelik kuralı yöntemi uygulanmıştır ve aşağıdaki kurallar oluşturulmuştur.

$$Destek = \frac{(Gofret, Yağ, Ekmek)}{Toplam\ Hareket} = \frac{2}{5} = 0{,}4$$

$$Güven = \frac{(Gofret, Yağ, Ekmek)}{(Gofret, Yağ)} = \frac{2}{4} = 0{,}5$$

{Gofret, Yağ} => Ekmek kuralında Destek değeri %40, güven değeri ise %50 olarak tespit edilmiştir. Bu sonuçları yorumlayacak olursak Destek değeri Gofret, Yağ, Ekmek ürünlerinden satın alan müşterilerin bütün alışverişlerdeki toplam hareket içerisindeki oranıdır. Güven değeri ise Gofret ve Yağ satın alan müşterilerin aynı alışveriş içerisinde Ekmek de satın alma oranını belirtmektedir. Çizelgeden elde edilen bazı kurallar şunlardır;

{Ekmek} →{Peynir} (Destek=0,6 Güven= 1)

{Ekmek, Peynir}→{Gofret} (Destek=0,6 Güven= 1)

{Peynir, Gofret} → {Yağ } (Destek=1 Güven= 0,8)

{Ekmek, Yağ} → {Peynir} (Destek=0,4 Güven= 1)

{Yağ} → {Ekmek} (Destek=0,8 Güven=0,5)

Birliktelik kuralında 2 önemli aşama vardır [9].

- Bütün nesne kümelerinin sık geçenler kümesinde yer alabilmesi için, her nesnenin daha önceden tanımlanmış olan eşik değeri yani minimum destek değerinden büyük olması gerekir.
- Sık geçen kümelerden güçlü ilişki kurallarının oluşturulmasında, belirlenen tanıma göre bu kurallar minimum destek ve minimum güven durumlarını sağlamalıdır.

Birliktelik kuralının matematiksel modeli Agrawal, Imielinski ve Swami tarafından ifade edilmiştir . Bu modele göre X= {y1,y2,y3……ym} nesnelerin kümesi ve D işlemler kümesi olarak ifade edilir. Her y ifadesi bir nesne olarak adlandırılır. D veri tabanında bütün işlemler T,T ⊆ X olacak şekilde tanımlanan nesnelerin kümesi(nesne küme) olsun. Her işlem bir tanımlayıcı alan olan TID ile temsil edilir. A ve B nesnelerin kümeleri olsun. Bir T işlemler kümesi ancak A⊆T ise yani A, T'nin alt kümesi ise A'yı kapsıyor denir. Bir birliktelik kuralı A ⇒ B formunda ifade edilir. A öncül(andecent) ve B izleyen (consequent) olarak adlandırılır. Burada, A ⊂ X, B ⊂ X ve A ∩ B = dır [8].

İlk olarak, A ⇒ B kuralı için d olasılığı ile kuralın destek değeri tanımlanır. Destek, T işleminin A ∪ B'yi içerme olasılığıdır. İkinci olarak, A ⇒ B kuralının g ile gösterilen güven değeri tanımlanır. Bu olasılık, T işleminin A'yı ve aynı zamanda B'yi içermesidir.

Matematiksel ifade ile kuralın destek ve güven değerleri aşağıdaki şekilde ifade edilir [8].

Destek (A ⇒ B) = P(A ∪ B)

Güven (A ⇒ B) = P(B \ A) veya

Güven (A ⇒ B) = Destek (A ⇒ B) / Destek (A)

Destek (A) = Destek (A ⇒ A)' dır.

Başka bir ifade ile destek ve güven değerleri aşağıdaki şekilde tanımlanır;

Destek (A) = │A│ / │D│

Destek (A ⇒ B) = │A.B│/ │D│

Güven (A ⇒ B) = Destek (A ⇒B) / Destek (A)

Burada; |A|, incelenen kayıtlardaki A ürünü içeren işlemlerin sayısı. |A.B|, incelenen kayıtlardaki A ve B ürünlerini birlikte içeren işlemlerin sayısı. |D|, veri tabanındaki bütün işlemlerin sayını ifade etmektedir.

3.2. Birliktelik Kuralı Algoritmaları

Günümüzde birliktelik kurallarının uygulanabilmesi için birçok algoritma geliştirmiştir. Bu algoritmalar oluşan ihtiyaçları karşılamak için kendi aralarında anlaşılabilirlik, kolay uygulanabilirlik ve performans vb.. olmak üzere çeşitli üstünlükler sağlayarak zaman içerisinde geliştirilmiştir. Birliktelik kurallarını oluşturmaya yönelik geliştirilen algoritmalardan bazıları aşağıda verilmiştir.

3.2.1. AIS algoritması

AIS algoritması, geniş nesne kümelerini üretmeyi amaçlayan bir algoritmadır.1993 yılında geliştirilmiştir. Çalışma mantığını inceleyecek olursak ilk olarak nesne adlarının A'dan Z'ye sıralanması ile başlar. Veri tabanını birçok kez tarar ve bütün işlemleri bu

esnada okur. İlk tarama esnasında veri tabanı içerisindeki boyut olarak geniş nesneler belirlenir. Bu belirlenen nesneler aday nesne kümeleri olarak tanımlanır. Bir işlem tarandıktan sonra, bir önceki taramada geniş oldukları belirlenen nesne kümeleriyle, o işlemin nesneleri arasındaki ortak nesne kümeleri belirlenir. Belirlenen bu ortak nesne kümeleri işlemde mevcut olan diğer nesnelerle birleştirilerek yeni aday kümeler oluşturulur. Herhangi bir nesne kümesi, bir işlemdeki nesnelerle birleşip aday kümelerden birini oluşturabilmesi için, birleşeceği nesnenin hem geniş olması hem de harf sırası açısından nesne kümesi içindeki tüm nesnelerden sonra geliyor olması gerekir.

AIS algoritması bu adımı gerçekleştirebilmek için, bir budama tekniği kullanır. Budama tekniğinin özünde, aday kümeler içindeki gereksiz kümelerin silinmesi vardır. Bu adımdan sonra, her aday kümesinin desteği hesaplanır. Destek seviyeleri minimum destek seviyesine eşit veya bu seviyeden büyük çıkanlar geniş nesne kümesi olarak işaretlenirler. Bir sonraki taramada bu geniş işareti taşıyan kümeler, yukarıda anlatıldığı gibi bir sonraki aday kümelerin belirlenmesi için kullanılır [19][55].

3.2.2. SETM algoritması

Bu algoritmada, Lk geniş nesne kümesinin her bir elemanı iki parametreden oluşur; bunların bir tanesi nesnenin ismiyken diğeri bu nesneyi ayırt etmeye yarayan bir özellik numarasıdır. Algoritma içinde bu numara TID (Transaction Identification) olarak kullanılır. Benzer şekilde her bir aday nesne kümeleri;

Ck,<TID, Nesne Kümesi İsmi>

formatında tutulur [56].

AIS'de olduğu gibi SETM algoritması da veri tabanını birçok kez tarar. İlk tarama esnasında veri tabanındaki nesneleri, teker teker sayarak hangilerinin geniş nesneler olduğunu belirler. Sonraki taramalarda, bir önceki geçişte geniş olarak işaretlenmiş nesne kümelerini kullanarak aday kümeleri belirler. Farklı olarak, SETM algoritması aday kümelerle birlikte üzerinde çalışılan işlemin TID bilgisini de tutar. Bundan sonra, aday nesne kümeleri nesne ismine göre sıraya dizilir ve küçük nesne kümeleri silinir. Eğer veri tabanı TID numarasına göre sıralanmışsa, bir sonraki tarama esnasında herhangi bir işlemdeki geniş nesne kümleri Lk'nın TID numarasına göre sıralanmasıyla elde edilir. Bu şekilde veri tabanı birkaç kez taranır. Artık başka herhangi bir geniş nesne kümesi bulunmadığında algoritma sonlandırılır.

SETM algoritmasında TID bilgisinin de tutulması, algoritmanın yer karmaşıklığını artıracaktır, bu dezavantajın dışında başka bir eksi nokta ise, aday nesne kümesinin desteği

hesaplanırken Ck sıralanmış halde değildir, bunun için nesne kümelerinin bir kez daha sıraya dizilmesi gerekecektir. Bu da zaman karmaşıklığını arttıran bir unsurdur [19].

3.2.3. Apriori algoritması

Apriori algoritması nesneler arasındaki bağlantılarının ortaya çıkarılmasında konusunda en çok bilinen ve en çok kullanılan algoritmadır [19]. Apriori Algoritmasının ismi, bilgileri bir önceki adımdan aldığı için "prior" anlamında Apriori'dir [53].

Geniş nesne kümelerini ortaya çıkartan algoritmalar eldeki tüm verileri birçok kez tararlar. İlk taramada, her bir nesnenin destek seviyesi, hesaplanarak kullanıcı tarafından başlangıçta girilen minimum destek seviyesi ile karşılaştırılır ve her bir nesnenin geniş olup olmadığına bakılır. Bundan sonraki her tarama bir önceki taramada geniş olarak tespit edilmiş nesnelerden başlar ve geniş nesne kümeleri oluşturulur. Bu geniş nesne kümelerine aday nesne kümeleri denir. Taramanın sonunda ise hangi aday nesne kümesinin gerçekten geniş olduğu kontrol edilir. Daha önce de belirtildiği gibi bir nesne kümesinin geniş olarak adlandırılabilmesi için o nesne kümesinin kullanıcı tarafından verilen minimum destek seviyesinin üzerine bir destek seviyesine sahip olması gerekir. Bir sonraki taramada, yine bir önceki taramada geniş olarak seçilen nesne kümelerinden başlanır ve veri tabanın sonuna kadar bu nesne kümelerinin destekleri hesaplanır. Bu işlem, başka yeni geniş nesne kümeleri bulunamayana kadar sürer [57].

Apriori algoritması daha önceden ortaya atılmış olan AIS ve SETM algoritmalarından her bir geçişte aday nesne kümelerinin sayılma ve bu aday kümelerinin üretilme şekliyle ayrılır. Hem AIS algoritmasında, hem de SETM algoritmasında, tarama esnasında, veriler okunurken aday nesne kümeleri üretilir. Bir işlem (T) (transaction) okunduktan sonra, geniş nesne kümelerinin bu işlemlerde olup olmadığına da bakılır. Yeni aday nesne kümelerinin üretilmesi ise işlemlerdeki diğer nesnelerle elde edilen geniş nesne kümelerinin birleştirilmesiyle üretilir [55]. Tabi bu da, gereksiz yere, aslında küçük nesne kümesi olan birçok aday nesne kümesinin sanki geniş nesne kümesiymiş gibi üretilmesi ve sayılması sonucunu doğurur. Bu da algoritmanın zaman karmaşıklığını arttırır.

Apriori algoritması ise aday nesneleri üretirken veri tabanındaki işlemleri hiç işin içine sokmadan, yalnızca bir önceki taramada geniş olduğu tespit edilmiş nesne kümelerini kullanarak oluşturur. Apriori algoritması geniş bir nesne kümesinin herhangi bir alt kümesinin de geniş olacağı kabulüne dayanır. Böylece k adet nesneden oluşmuş bir nesne kümesi, k-l adet nesneye sahip geniş nesne kümelerinin birleştirilmesi ve alt kümeleri geniş olmayanların silinmesiyle elde edilebilir. Bu

birleşme ve silme işlemi sonunda daha az sayıda aday nesne kümeleri oluşacaktır [19]. Agrawal ve Srikant algoritmanın çalışma ayrıntılarını ve algoritmanın kaba kodunu şu şekilde ifade etmiştir [53];

- Verilerin ilk taranması esnasında, geniş nesne kümelerinin tespiti için, tüm nesneler sayılır.
- Bir sonraki tarama, k'ıncı tarama olsun, iki aşamadan oluşur;
- Apriori-gen fonksiyonu kullanılarak, (k-1) inci taramada elde edilen, Lk-1 nesne kümeleriyle, Ck aday nesne kümeleri oluşturulur,
- Sonra veri tabanı taranarak, Ck daki adayların desteği sayılır.
- Hızlı bir sayım için, verilen bir 1 işlemindeki, Ck'yı oluşturan adayların çok iyi belirlenmesi gerekir.

Apriori Algoritmasının klasik özet kodu Şekil 3.1.'de görülmektedir [F19].

```
L1= {sık geçen 1-öğe kümesi};
  for (k=2; Lk-1≠ Ø; k++) do begin
      Ck=apriori-gen (Lk-1); // Yeni adaylar
      forall transactions-hareketler t ∈ D do begin
          Ct = subset (Ck, t); // Adaylar t içindedir
          forall candidates – adaylar c ∈ Ct do
              c.count++;
      end
      Lk = {c ∈ Ck | c.count ≥ minsup}
  end
Answer = ∪k Lk;
```

Şekil 3.1. Klasik Apriori Algoritması Özet Kodu

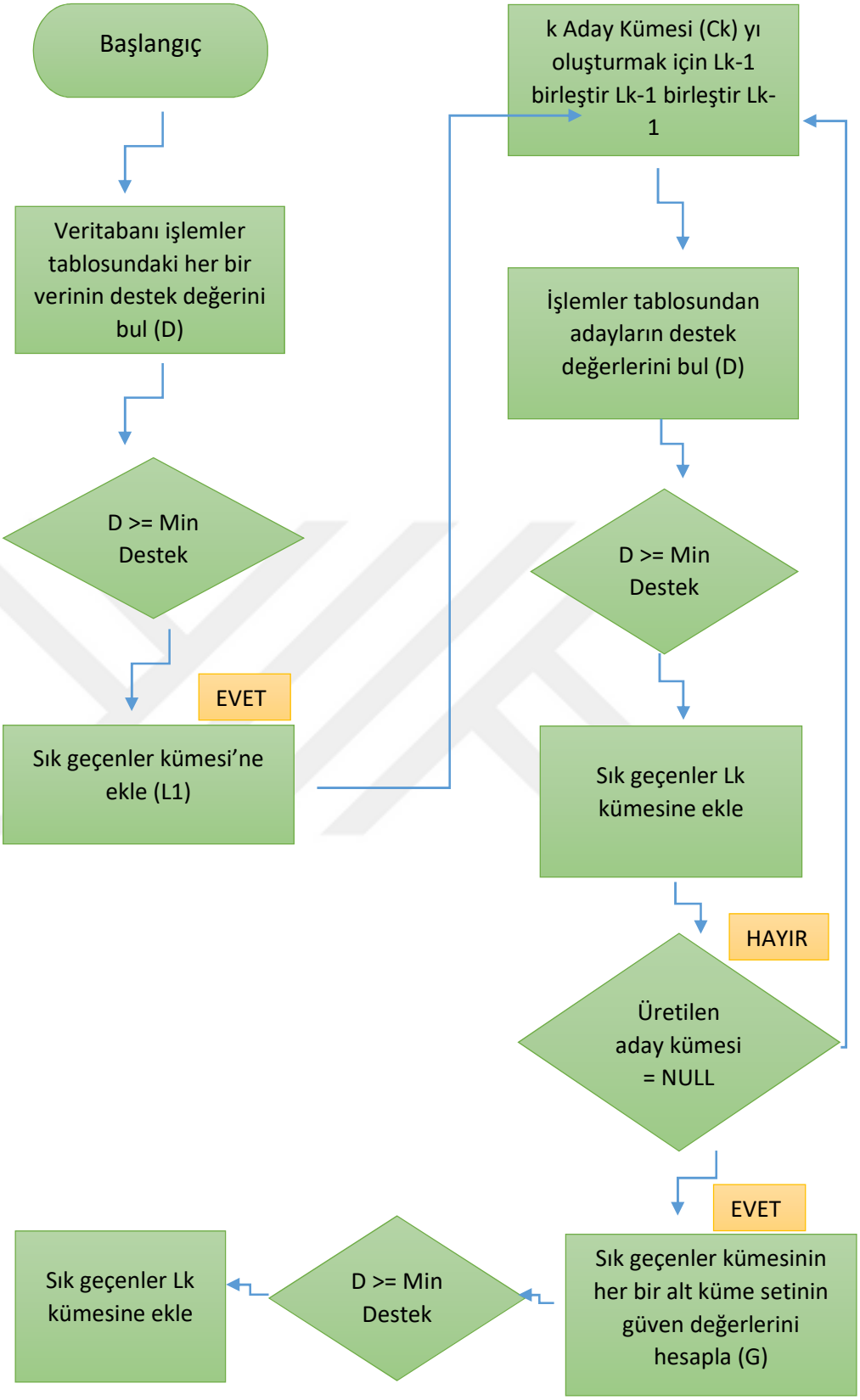
Apriori algoritmasının akış diyagramı aşağıdaki Şekil 3.2. de belirtilmiştir.

Apriori Algoritması Örnek Uygulama

Apriori algoritmasının daha kolay anlaşılabilmesi için ana uygulama öncesinde bir örnek çalışma ile ayrıntılı olarak verilmiştir. Örnek olarak 5 müşterinin markette yaptıkları alışverişleri göz önüne alalım. Çizelge 3.2. de bu müşterilere ait alışveriş verileri bulunmaktadır. Her satır, ilgili müşterinin tek seferde yaptıkları ürün tercihlerini göstermektedir.

Çizelge 3.2. Müşteri alışverişleri

Müşteri No	Satın Alınan Ürünler
1	Tuz, Limon, Zeytin
2	Zeytin, Un, Süt, Tereyağı
3	Tuz, Un, Çamaşır Suyu, Zeytin, Tereyağı
4	Zeytin, Un, Limon, Tereyağı
5	Un, Tereyağı, Tuz, Ayran



Şekil 3.2. Apriori Algoritması Akış Diyagramı

Bu çizelgedeki veriler kullanılarak müşterilerin satın alma alışkanlıklarının belirlenmesi hedeflenmiştir. Çözümlemeden önce destek(support) ve güven (confidence) değerleri için eşik değerleri belirlenmelidir. Bu eşik değerleri doğrultusunda algoritmanın işleyişi şekillenecektir.

Destek değeri = % 60 - Güven değeri = %75

Buradaki destek(eşik) = %60 değeri tüm müşteri sayısının 10 olduğunu varsayarsak destek sayısının 0.6*10 = 6 olduğu anlaşılır. Beş müşterinin satın almış oldukları ürünlerin kümesi {Tuz, Limon, Zeytin, Un, Süt, Tereyağı, Çamaşır Suyu, Ayran} dır. Bu ürünlerin hepsi için destek değerleri hazırlanacaktır. Çizelge 3.2. deki müşteri hareketleri incelendiği zaman Tuz ürününden 1, 3 ve 5. Müşteriler satın almıştır. Bu durumda sayı(Tuz) ifadesi 3 olmaktadır. Benzer şekilde değer ürünler içinde destek değerleri hesaplanır. Çizelge 3.3 de bütün ürünlerin destek değerleri verilmektedir.

Çizelge 3.3. Tekil ürünlerin tercih edilme sayısı

Ürün	**Tercih Edilme Sayısı**
Tuz	3
Limon	2
Zeytin	4
Tereyağı	4
Un	4
Çamaşır Suyu	1
Ayran	1
Süt	1

Bu çizelgedeki ürünlerden eşik değeri yani sayısı 3'ün altında olan ürünleri çıkarıyoruz. Geriye kalan (Tuz, Zeytin, Tereyağı, Un) ürünlerini ikili ürün grupları olacak şekilde hazırlayıp aşağıdaki gibi destek sayılarını tespit ediyoruz. Çizelge 3.4 de bu destek sayıları verilmiştir.

Çizelge 3.4. İkili ürün grubu destek sayısı

Sayı(Tuz, Zeytin) =2

Sayı(Tuz, Tereyağı) = 2

Sayı(Tuz, Un) = 2

Sayı(Zeytin, Tereyağı) = 3

Sayı(Zeytin, Un) = 3

Sayı(Tereyağı, Un) = 4

Ürün	**İkili Ürün Grubu Adet**
Tuz, Zeytin	2
Tuz, Tereyağı	2
Tuz, Un	2
Zeytin, Tereyağı	3
Zeytin, Un	3
Tereyağı, Un	4

Bu adımda da aynı şekilde eşik değeri 3 ten küçük olan ürünleri çizelgeden çıkaracağız. Geriye aşağıdaki Çizelge 3.5 deki ürün grupları kalacaktır.

Çizelge 3.5. Eşik değerinin üstünde kalan ürün grupları

Ürün grupları	**Tercih edilme sayısı**
Zeytin, Tereyağı	3
Zeytin, Un	3
Tereyağı, Un	4

Çözümlemeye katılacak ürünler tespit edildikten sonra Çizelge 3.2 den faydalanacak üçlü ürün grupları oluşturularak, destek sayıları hesaplanır. Bu destek sayıları Çizelge 3.6. da verilmiştir.

Çizelge 3.6. Üçlü ürün grupları destek sayıları

Ürün Grupları	**Tercih edilme sayısı**
Zeytin, Tereyağı, Tuz	1
Zeytin, Tereyağı, Limon	1
Zeytin, Tereyağı, Un	3
Zeytin, Tereyağı, Çamaşır Suyu	1
Zeytin, Tereyağı, Ayran	0
Zeytin, Tereyağı, Süt	1
Zeytin, Un, Tuz	1
Zeytin, Un, Limon	1
Zeytin, Un, Çamaşır Suyu	1
Zeytin, Un, Ayran	0
Zeytin, Un, Süt	1
Tereyağı, Un, Tuz	2
Tereyağı, Un, Limon	1
Tereyağı, Un, Çamaşır Suyu	1
Tereyağı, Un, Ayran	1
Tereyağı, Un, Süt	1

Eşik değeri 3'ün altında olan ürünleri çözümlemeden çıkarıyoruz. Geriye eşik değerini sağlayan tek üçlü ürün grubu olduğu görülüyor. Bu son adımda birliktelik kuralını elde edebiliriz. (Zeytin, Tereyağı, Un) ürünleri müşteriler tarafından aynı alışveriş içerisinde 3 defa tercih edilmiştir. Eşik değerini sağlayan bu ürün grubuna ait destek ve güven değerleri hesaplanacaktır.

- {Zeytin, Tereyağı, Un} ürün kümesi göz önüne alındığında güven değeri aşağıdaki şekilde elde edilir.

$$Güven(Zeytin, Tereyağı \rightarrow Un) = \frac{Sayı\{Zeytin, Tereyağı, Un\}}{Sayı\{Zeytin, Tereyağı\}} = \frac{3}{3} = \%100$$

- {Zeytin -- > Un, Tereyağı} birliktelik kuralı için güven ölçütü şu şekilde hesaplanır.

$$\text{Güven}(\text{Zeytin} \rightarrow \text{Un}, \text{Tereyağı}) = \frac{Sayı\{Zeytin, Tereyağı, Un\}}{Sayı\{Zeytin\}} = \frac{3}{4} = \%75$$

- {Un -- > Zeytin, Tereyağı} birliktelik kuralı için güven ölçütü şu şekilde elde edilir.

$$Güven(Un \rightarrow Zeytin, Tereyağı) = \frac{Sayı\{Zeytin, Tereyağı, Un\}}{Sayı\{Un\}} = \frac{3}{4} = \%75$$

- {Tereyağı -- > Zeytin, Un} birliktelik kuralı için güven ölçüsü aşağıdaki şekilde hesaplanır.

$$\text{Güven(Tereyağı -- > Zeytin, Un)} = \frac{Sayı\{Zeytin, Tereyağı, Un\}}{Sayı\{Tereyağı\}} = \frac{3}{4} = \%75$$

Elde edilen tüm güven ölçüleri başlangıçta ilan edilmiş olan güven eşik değerinden büyüktür. Sonuç olarak aşağıdaki Çizelge 3.7. deki birliktelik kuralları elde edilmiştir.

Çizelge 3.7 Birliktelik kuralları

Birliktelik Kuralı	Anlamı	Güven Oranı
Zeytin & Tereyağı →Un	Zeytin ve Tereyağının bulunduğu ürün kümesinde Un ürününün olma olasılığı	%100
Zeytin → Un & Tereyağı	Zeytinin yer aldığı ürün kümesinde Un'un ve Tereyağının bulunma olasılığı	%75
Un → Zeytin & Tereyağı	Un'un yer aldığı bir ürün kümesinde Zeytin ve Tereyağı bulunma olasılığı	%75
Tereyağı → Zeytin & Un	Tereyağının yer aldığı bir ürün kümesinde Zeytin ve Un'un buluma olasılığı	%75

3.2.4 Apriori TID algoritması

Önceki bölümlerde verilen algoritmalarda destek değerini hesaplamak için bütün veri tabanı taranır; ancak her aşamada taranmaması gereken durumlar olabilir. Bu yaklaşımla Agrawal, Apriori algoritmasıyla birlikte AprioriTid algoritmasını da sunmuştur. AprioriTid algoritması da taramadan önce aday nesne kümelerini belirlemek için Şekil 3.3. de görülen apriori-gen fonksiyonu kullanılmaktadır. Apriori'den en büyük farkı ilk geçişten sonra veri tabanının destek seviyesini bulmak için taranmamasıdır. Bu iş için Ck kullanılır. SETM algoritmasında olduğu gibi Ck'nin her elemanı <TID,{Xk}> formundadır. Burada Xk, TID numaralı işlemde bulunan potansiyel geniş k nesne kümesidir. k= 1 iken C1 veri tabanına karşılık gelir. Bununla beraber her nesne adımında olduğu gibi üretilir. t işlemindeki Ck bir elemanı <TID,c> şeklindedir. Burada c, t işlemindeki Ck'ya ait bir aday elemandır. {c∈Ck|c}, Eğer bir işlemin, herhangi bir k nesne kümesi adayı yoksa, bu durumda Ck nın bu işlem için herhangi bir girdisi, elemanı olmayacaktır. Daha doğrusu bu işlemin TID numarasını taşımıyor olacaktır. Böylece Ck daki girdi sayısı, özellikle bu k değerleri için, veri tabanındaki işlem sayısından daha küçük olabilir. Bunun dışında yine büyük k değerleri için her girdi kendisine karşılık gelen işlemlerden daha küçük olabilir.

Çünkü o işlemde çok az sayıdan aday barınıyor olabilir. Ancak, küçük k değerleri için bunun tersi olacaktır; yani girdiler kendilerine karşılık gelen işlemlerden daha büyük olabilecektir [53]. Agrawal ve Srikant bildiride, AprioriTid algoritmasının kaba kodunu şu şekilde açıklamışlardır.

```
L1 = {Geniş 1-nesne kümeleri} ;
C1 = D veritabanı;
for (k=2; Lk-1≠ 0; k++) do begin
    Ck=apriori-gen (Lk-1); // Yeni adaylar
    Ck = 0;
        for all işlemler t ∈ Ck-1 do begin
// TID numaralı işlemlerdeki Ck-1 içinde bulunan tüm aday nesne kümelerini belirle
        Ct = { c ∈ Ck|(c-c[k]) ∈ t.nesnekümeleri ^ (c-c[k-1]) ∈ t.nesnekümeleri }
                    for all adaylar c ∈ Ct do
                    c.count ++;
                    if (Ct ≠ 0) then Ck += <t.TID, Ct >
    end
    Lk = { c ∈ Ck | c.count >= min-destek}
end
Yanıt = Uk Lk
```

Şekil 3.3. AprioriTid algoritması sözde kodu[F53]

3.3 Birliktelik Kuralı Kullanım Alanları

Birliktelik kuralı yöntemleri birçok sektörde çeşitli analiz işlemlerinin bir parçası olarak aktif olarak kullanılmaktadır [57]. 2000' den 2014 yılına kadar olan dönemde birliktelik kuralı analiz çalışmalarının nasıl değiştiği literatür taraması ile incelenmiştir [57]. Çizelge 3.8. de son yıllarda birliktelik kuralları ile alakalı yapılan uluslararası yayınlar verilmiştir.

Çizelge 3.8. 2000-2014 yılları arasındaki birliktelik kurallarına yönelik yapılan makale sayısı [57].

Yıllar	Makale Sayısı
2000	0
2001	0
2002	0
2003	1
2004	0
2005	0
2006	0
2007	2
2008	1
2009	3
2010	0
2011	1
2012	2
2013	3
2014	2

4. MÜŞTERİ PROFİLİNE DAYALI MARKET SEPET ANALİZİ UYGULAMASI

Bu bölümde daha önceki bölümlerde bahsedilmiş olan müşteri sepet analizi ve müşteri kartlarına ait tanımlamaları destekleyecek örnek bir çalışma yapılmıştır. Yüksek boyutlu gerçek satış verisi kullanılarak yapılacak bu çalışmada, müşterilerin kişisel özelliklerine göre tercih ettikleri ürünlerin birliktelik analizi yapılmıştır. Çalışmada yapay sinir ağları teknikleri kullanılarak yapılan tahmin işlemleri sayesinde birliktelik kuralları kullanılarak oluşturulabilecek kampanyaların verimliliğini önceden tespit edilmesi sağlanmıştır. Bu çalışmanın ilk aşamasında analiz edilecek olan verinin özelliklerinden bahsedilmiştir. Veriler analiz için uygun formata getirilmiş ve ardından çıkarılan birliktelik kuralları müşterilere ait bilgiler ile bağlantılı hale getirilerek müşterilerin tercihleri ve bu tercihlerin müşterilerin özellikleri üzerine dağılımı ortaya çıkarılmıştır. Daha sonraki aşamalarda bu verilere ait analiz sonuçları daha anlaşılabilir ve yorumlanabilir olması için web tabanlı bir program üzerinden görsel olarak veriler sunulmuştur. Hazırlanan dinamik web tabanlı uygulama üzerinden müşteri- ürün ilişkisi gözetilerek yapılacak kampanyalar için önemli bilgilerin elde edilmesi sağlanmıştır. Bu analiz çalışmasının 2 temel süreci vardır. Bunlarda birincisi R programlama dili ile marketlerdeki satışlara ait en yüksek destek ve güven değerine sahip birliktelik kurallarının tespiti, diğeri ise sqlserver üzerinden sql komutları ile bu belirlenen birliktelik kurallarını temsil eden müşteri kitlesinin ortaya çıkarılmasıdır.

4.1. Analizde Kullanılacak Veri Seti

Analiz işleminde kullanılacak veri birçok şubesi bulunan bir süpermarketler zincirinin 2 yıla ait satış bilgilerini kapsamaktadır. Marketler bulundukları şehirde homojen olarak hemen hemen bütün mahallelerde hizmet verecek şekilde dağıldığı için bölgesel olarak yapılacak analiz işlemlerinin doğruluk payını arttırdığı söylenebilir. Analiz işlemi büyük oranda 1. yıldaki veriler üzerinden gerçekleştirilmiştir. Fakat bazı karşılaştırma işlemleri için 2. yıla ait verilerinin de kullanıldığı alanlar olmuştur. Analiz işlemlerinde şirketin gizlilik politikası gereği müşterilerin ait bazı bilgilerin gösterimi kısıtlanmıştır. Veriler ile ilgili bazı istatistiksel bilgiler Çizelge 4.1. de belirtilmiştir.

Çizelge 4.1. Veri istatistikleri

Veri Başlığı	Birinci Yıl(Adet)	İkinci Yıl(Adet)
Toplam Satış Hareket Kaydı	43.569.223	42.600.584
Toplam Alışveriş Sayısı	10.263.619	10.321.732
Müşteri Kartı İle Toplam Satış Hareket Kaydı	8.272.668	5.661.958
Müşteri Kartı İle Yapılan Toplam Alışveriş Sayısı	954.415	655.248
Müşterilerin Doğum Tarihinin Bulunduğu Alışveriş Sayısı	922.216	629.319

Cinsiyet Bilgisinin Bulunduğu Alışveriş Sayısı	954.415	655.248
Öğrenim Durumu Bulunduğu Alışveriş Sayısı	454.341	308.765
Adres Bilgisinin Bulunduğu Alışveriş Sayısı	625.715	442.547
Meslek Bilgisinin Bulunduğu Alışveriş Sayısı	365.839	250.676
Medeni Hâli Bilgisinin Bulunduğu Alışveriş Sayısı	954.415	655.248

Bu iki yıldaki toplam veri yaklaşık olarak 86 milyon satırdır. Bu kadar yüksek boyuttaki veriyi işlemek ve kısa sürede yorumlanabilir hale gelmesi için performanslı bir veri tabanı yönetim sistemi gerekmektedir. Bu tez çalışmamızda MSSQL Server 2014 Developer Edition sürümü kullanılmıştır. Çizelge 4.1. de verilen istatistikler MSSQL Server veri tabanında gerekli filtrelemeler yapılarak elde edilmiştir.

Şekil 4.1'de veri tabanı üzerindeki ürünlerin satış hareketleri ile ilgili verilerin bir kısmı verilmiştir.

FİŞ KODU	ALIŞVERİŞ YAPILAN MARKET	ÜRÜN KODU	ÜRÜN ADI	SATILAN MİKTAR	ÜRÜN FİYATI	ÜRÜNÜN TOPLAM TUTARI	ALIŞVERİŞ TARİHİ
10420000033104	1042	8690536012312	TENO PECETE 200 LU *16*	1	1,69	1,69	2014-12-31 19:49:20.000
10420000033104	1042	8690536012312	TENO PECETE 200 LU *16*	1	1,69	1,69	2014-12-31 19:49:20.000
10420000033105	1042	8690574950515	PEPSI 2.5 LT JUMBO *6*	1	3,35	3,35	2014-12-31 19:49:52.000
10420000033105	1042	8690504034902	ULKER 349-00 METRO 5 LI 200 GR *18*	1	2,20	2,20	2014-12-31 19:49:52.000
10420000033105	1042	2600215	URF.20.KARISIK CEREZ 1 KG	0,274	21,90	6,00	2014-12-31 19:49:52.000
10420000033106	1042	6082001177510	ISTER EKMEK	3	0,70	2,10	2014-12-31 19:54:54.000
10420000033107	1042	8690524000246	FALIM SAKIZ 5'LI ORMAN MEYVELI 7 GR *20*	1	0,45	0,45	2014-12-31 20:01:06.000
10420000033107	1042	8690524000246	FALIM SAKIZ 5'LI ORMAN MEYVELI 7 GR *20*	1	0,45	0,45	2014-12-31 20:01:06.000
10420000033108	1042	8690533084732	ETI BALIK KRAKER MISIRLI 95 GR*10*	1	0,95	0,95	2014-12-31 20:04:59.000
10420000033108	1042	8690533084732	ETI BALIK KRAKER MISIRLI 95 GR*10*	1	0,95	0,95	2014-12-31 20:04:59.000
10420000033108	1042	8690767675089	SUTAS PEYNIR 500 GR SUZME*24*	1	7,25	7,25	2014-12-31 20:04:59.000
10420000033108	1042	8692971031316	ULKER ICIM KASAR 700 GR *10*	1	13,90	13,90	2014-12-31 20:04:59.000
10420000033108	1042	8691375226717	BIZIM MISIROZU 2 LT.*9*	1	11,99	11,99	2014-12-31 20:04:59.000
10420000033108	1042	2600120	KESTANE	0,706	14,20	10,03	2014-12-31 20:04:59.000
10420000033108	1042	2600059	KARNIBAHAR	1,566	1,45	2,27	2014-12-31 20:04:59.000
10420000033108	1042	8690504089971	ULKER 383-6 9 KAT MUZLU GOF.84 GR *18*	1	1,10	1,10	2014-12-31 20:04:59.000
10420000033108	1042	8690504089971	ULKER 383-6 9 KAT MUZLU GOF.84 GR *18*	1	1,10	1,10	2014-12-31 20:04:59.000
10420000033108	1042	8690560051189	BAG.MISIR UNU EKO POS.250 GR*10*	1	1,99	1,99	2014-12-31 20:04:59.000
10420000033108	1042	8690504025573	ULKER 255-7 A.FIST.BIT.80 GR*36*	1	3,00	3,00	2014-12-31 20:04:59.000
10420000033108	1042	8690526063140	ETI CIN COKLU 377 GR *20*	1	2,49	2,49	2014-12-31 20:04:59.000
10420000033108	1042	8690526083292	ETI BENIMO LOKMALIK 77 GR *24*	1	0,95	0,95	2014-12-31 20:04:59.000
10420000033108	1042	8690526083292	ETI BENIMO LOKMALIK 77 GR *24*	1	0,95	0,95	2014-12-31 20:04:59.000
10420000033108	1042	8699872790014	NESIL YUFKA	1	4,25	4,25	2014-12-31 20:04:59.000

Şekil 4.1 Ürün satış hareketleri

Veri tabanında kayıtlı bulunan bütün ürünler gruplandırılmıştır. Veri tabanında 13643 farklı ürüne ait stok hareketi bulunmaktadır. Çizelge 4.2. de sınıflandırmada kullanılan ürünlere ait kaç farklı grup olduğuna dair bilgiler verilmiştir. Ayrıca Şekil 4.2'de Analiz işleminde kullanılacak ürünlerin bağlı bulundukları 5 kırılımlı hiyerarşik gruplar verilmektedir. Analiz çalışmamızda çıkarılacak birliktelik kuralı Şekil 4.2. de belirtildiği üzere Kod_3 ürün grubuna göre yapılmıştır.

Çizelge 4.2. Birbirinden farklı ürün ve ürün grubu sayıları

Grup İsimleri	Farklı ürün ve grup sayısı
Ürünler	13.643
Ana Grup	10
Kod_1	43
Kod_2	184
Kod_3	602
Kod_4	789

ÜRÜN KODU	ÜRÜN ADI	ANAGR...	KOD_1	KOD_2	KOD_3	KOD_4
2600669	NERGIS YUFKA KG	GIDA	SÜT-KAHVALTILIK	PASTANE ÜRÜNLERİ	YUFKA	YUFKA
8698849880017	YUMURTA MALYUM 30'LU SARI E.ANAC	GIDA	SÜT-KAHVALTILIK	YUMURTA	NORMAL YUMURTALAR	NORMAL YUMURTALAR
8699971794166	PAREX STRECH FILM 15 MT*24*	EV-HOBI	MUTFAK EŞYA-GEREÇLERİ	MUTFAK SARF MALZEMELERİ	STREÇ FİLM	STREÇ FİLM
8690637582721	CARTE D'OR PUD.BITTER CIK.116 GR.*12*	GIDA	GIDA-ŞEKERLEME	UNLU MAMÜL-TATLI	TOZ TATLILAR	PUDİNG
8690637582721	CARTE D'OR PUD.BITTER CIK.116 GR.*12*	GIDA	GIDA-ŞEKERLEME	UNLU MAMÜL-TATLI	TOZ TATLILAR	PUDİNG
2600253	ERCIYES VAKUMLU MARAS SUCUGU	GIDA	GIDA-ŞEKERLEME	SAKIZ-ŞEKERLEME	ŞEKERLEMELER	GELENEKSEL ŞEKERLEME
8692095337738	SEK MEYVE SUYU 1/5 KARISIK *27*	GIDA	İÇECEK	GAZSIZ İÇECEK	MEYVE SUYU	MEYVE SULARI
8680035270108	AK BINSAN CUBUK SEKER	GIDA	GIDA-ŞEKERLEME	SAKIZ-ŞEKERLEME	ŞEKERLEMELER	YUMUŞAK ŞEKER
8690632055381	NESTLE 1927 EXT.SUTLU CIK. 80 GR *144*	GIDA	GIDA-ŞEKERLEME	ÇİKOLATA-GOFRET	ÇİKOLATA	TABLET ÇİKOLATA
8690632055381	NESTLE 1927 EXT.SUTLU CIK. 80 GR *144*	GIDA	GIDA-ŞEKERLEME	ÇİKOLATA-GOFRET	ÇİKOLATA	TABLET ÇİKOLATA
8693371010215	KARLIDAG SUT YARIM YAGLI 1/1 *12*	GIDA	SÜT-KAHVALTILIK	SÜT	UZUN ÖMÜRLÜ SÜT	UZUN ÖMÜRLÜ SÜT
2600253	ERCIYES VAKUMLU MARAS SUCUGU	GIDA	GIDA-ŞEKERLEME	SAKIZ-ŞEKERLEME	ŞEKERLEMELER	GELENEKSEL ŞEKERLEME
8690547030022	KENTON 5'LI VANILIN *180*	GIDA	GIDA-ŞEKERLEME	UNLU MAMÜL-TATLI	PASTA MALZEMELERİ	KABARTMA TOZU VE ŞEKERLİ VANİLİN
2600387	DANA KUSBASI	GIDA	ET-BALIK	KIRMIZI ET	AÇIK DANA ETİ	KUŞBAŞI ET
8690547030022	KENTON 5'LI VANILIN *180*	GIDA	GIDA-ŞEKERLEME	UNLU MAMÜL-TATLI	PASTA MALZEMELERİ	KABARTMA TOZU VE ŞEKERLİ VANİLİN
8690710137596	KOSKA 80 GR CEVIZLI SUCUK	GIDA	GIDA-ŞEKERLEME	SAKIZ-ŞEKERLEME	ŞEKERLEMELER	GELENEKSEL ŞEKERLEME
8690710137596	KOSKA 80 GR CEVIZLI SUCUK	GIDA	GIDA-ŞEKERLEME	SAKIZ-ŞEKERLEME	ŞEKERLEMELER	GELENEKSEL ŞEKERLEME
8692971416373	ULKER SMART CIK. SUT 200 ML *27*	GIDA	SÜT-KAHVALTILIK	SÜT	UZUN ÖMÜRLÜ SÜT	UZUN ÖMÜRLÜ SÜT
8692971416588	ULKER SMART MUZLU SUT 200 ML *27*	GIDA	SÜT-KAHVALTILIK	SÜT	UZUN ÖMÜRLÜ SÜT	UZUN ÖMÜRLÜ SÜT
86914970	NESQUIK SUT KAK.180 ML*27*	GIDA	SÜT-KAHVALTILIK	SÜT	UZUN ÖMÜRLÜ SÜT	UZUN ÖMÜRLÜ SÜT

Şekil 4.2. Ürünlerin hiyerarşik alt grupları

Şekil 4.3.'de Veri tabanındaki müşterilerin kişisel bilgilerine ait birkaç örnek verilmiştir. Müşterilerin Soyisim ve Adres bilgileri şirketin gizlilik politikası gereğince gösterilmemiştir. Müşteriler ile ilgili sahip olduğumuz veriler müşteri adı, medeni hâli, cinsiyet, mesleği, öğrenim durumu, adres ve müşterinin doğum tarihinden oluşmaktadır.

MÜŞTERİ KART NO	MÜŞTERİ İSMİ	MEDENİ HALİ	CINSIYET	MESLEĞİ	ÖĞRENİM DURUMU	ADRESİ	DOĞUM TARİHİ
20100000	HACIBEKİR	0	ERKEK	SERBEST MESLEK	LİSE	BEYDA	1966
20100000	MEHMET	1	ERKEK	ÖĞRENCİ	ÜNİVERSİTE	DOLAM	1987
20100000	SEMRA	0	KADIN	ÖĞRETMEN	ÜNİVERSİTE	BORAN	1971
20100000	ALİ	0	ERKEK	İŞÇİ	LİSE	BATTAL	1956
20100000	YETER	1	KADIN	SERBEST MESLEK	İLKÖĞRETİM	SİVAS C	1980
20100000	NESRİN	0	KADIN	EV HANIMI	LİSE	İNÖNÜ	1962
20100000	SEHER	1	KADIN	ÖĞRETMEN	ÜNİVERSİTE	DERNE	1987
20100000	ASUMAN	0	KADIN	EV HANIMI	İLKÖĞRETİM	ÖZALPE	1940
20100000	NURDAN	1	KADIN	MUHASEBECİ	ÜNİVERSİTE	CENGİZ	1979
20100000	NURSEL	0	KADIN	EV HANIMI	İLKÖĞRETİM	2. CAD.	1965
20100000	SELÇUK	1	ERKEK	İŞÇİ	ÜNİVERSİTE	KERNE	1984
20100000	TÜRKAN	0	KADIN	SERBEST MESLEK	LİSE	FIRAT M	1978
20100000	NURHAN	0	KADIN	EV HANIMI	LİSE	YAKINC	1968
20100000	HÜLYA	0	KADIN	EV HANIMI	LİSE	UÇBAĞ	1973
20100000	HASAN	0	ERKEK	SERBEST MESLEK	İLKÖĞRETİM	CEVHE	1970
20100000	OSMAN	0	ERKEK	ÖĞRETİM ÜYESİ	YÜKSEK LİSANS	KERNE	1963
20100000	MAHMUT	0	ERKEK	işçi	ÜNİVERSİTE	2.ORDU	1957
20100000	SELAHATTİN	0	ERKEK	ÇİFTÇİ	İLKÖĞRETİM	SALKÖ	1964
20100000	NURETTİN	0	ERKEK	ÇİFTÇİ	İLKÖĞRETİM	2. CAD.	1961

Şekil 4.3 Müşteri bilgileri

4.2. Müşteri Kartları

Perakende sektöründeki en önemli başarı faktörü sürekliliktir. Sürekli olarak artan müşteri profilinin oluşturulmasının yanı sıra var olan müşterilerin sadakatini kazanma da önemli bir başarı olarak nitelendirilebilir. Nakit akışının, satın alma ve karlılığın sürekli olması için ürün satışlarının belirli bir düzende ve sürekli olması gerekmektedir. Bu süreç de müşterilerin sürekliliğini sağlamaktan geçmektedir. Müşteri kartları tamda bu noktada devreye girmektedir. Bu kartlar sayesinde müşterilere ait bir çok bilgi sisteme entegre edilip, alışveriş işlemi sırasında müşterilerin alışveriş eğilimlerini tespit etme imkanı sağlamaktadır. Müşteri kartları sayesinde müşteriler bireysel olarak değerlendirilebildiği gibi müşteriler gruplandırılarak kitlesel olarak da değerlendirilebilmektedir. Birkaç örnek verilecek olursa; Müşteri hangi aylarda en çok hangi ürünleri alıyor, indirim dönemlerinde alışverişe ne kadar duyarlı, alışverişlerini daha çok hangi günlerde yapıyor, ortalama alışveriş tutarı, 30-40 yaşları arasındaki müşterilerin en çok satın aldıkları ürünler veya belirli bir bölgede oturan müşterilerin tercih ettikleri ürün grupları ve ürünler olarak tespit edilebilmektedir. Ayrıca sepet analizi gibi veri madenciliği teknikleri ile tespit edilen veriler ışığında müşterilerin satın alma eğilimlerinin ortaya çıkarılmasında da müşteri kartları önemli bir rol oynamaktadır. Birçok marka, elde edilen bu analiz sonuçları ile müşteri grubu ve bireysel müşteri odaklı kampanyalar yaparak müşterilerin sadakat oranını artırmak istemektedir. Müşteri kartlarının hem firmaya hem de müşterilere birçok faydası mevcuttur. Bu faydalardan bazıları;

Müşteri kartlarının firmaya sağladığı faydalar;

- Müşterilerini kayıt altına alarak potansiyel müşteri profili'ni oluşturma
- Müşterilere özel kampanyalar hazırlayarak ürün satışını artırma
- Müşterilerin bilgilerini kullanarak yapılan kampanyaları direk ilgili müşteri gruplarına ulaştırma imkânına sahip olma

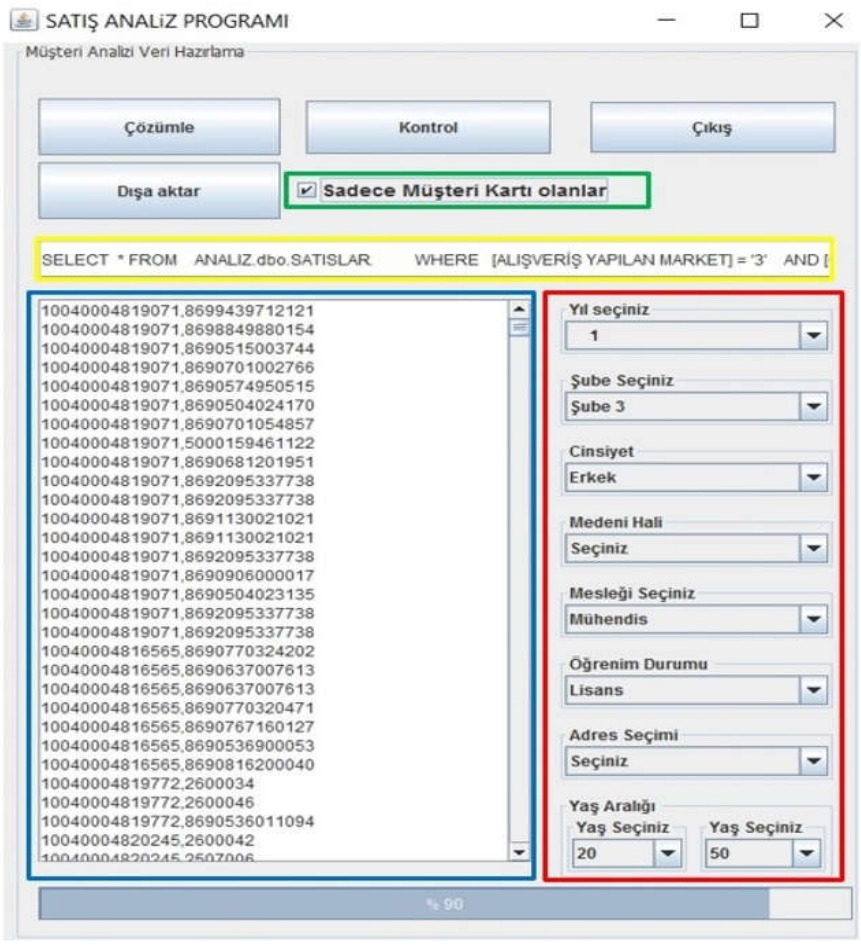
Müşteri kartlarının müşterilere sağladığı faydalar;

- Anında indirim yapılması
- Aldıkları belirli ürünler karşılığında müşterilere puan verilmesi
- Aldıkları puan karşılığı alışveriş çeki verilmesi ve özel hediyeler verilmesi
- Mağazanın yaptığı kampanyalara katılma hakkı verilmesi
- Aldıkları puan karşılığında mağazadaki ürünlerden bedava verilmesi
- Mağazanın anlaşmalı olduğu diğer firmalardan indirim verilmesi

4.3. Veri Hazırlama Programı

Verilerin içeriğinden önceki bölümlerde bahsedilmiştir. R-Studio'da veriler için apriori algoritmasının uygulanabilmesi için belirli bir formatta olması gerekmektedir. Veritabanın'daki verilerin hızlı bir şekilde uygun formata getirilmesi ve gerektiğinde verileri filtrelemek için bir program hazırlanmıştır. Şekil 4.4. de görüldüğü üzere program kırmızı çerçeve içerisindeki combobox'lar ile seçilmiş özelliklere uygun olan satış hareketlerini getirmektedir. Arka planda çalıştırdığı sql sorgusunun bir kısmı ise programda sarı çerçeve içerisinde belirtilmiştir. Analiz işlemine tabi tutulacak veriler ise mavi çerçeve içerisinde gösterilmiştir. Ayrıca müşteri kartı ile yapılan alışveriş hareketleri tercihe bağlı olarak değiştirilebilmesi için yine programa yeşil çerçevedeki seçim butonu eklenmiştir. Aktif olması durumunda program sadece müşteri kartları ile yapılan alışverişleri göz önene alacaktır.

Bir örnek ile açıklayacak olursak Şekil 4.4. deki programın yaptığı filtrelemeler sonucunda 1. yıla ait 3 numaralı şubede 20 ile 50 yaşları arasındaki mesleği mühendislik olan lisans mezunu erkek müşterilerin satış hareketlerini getirmektedir. Program bize verinin daha rahat bir şekilde hazırlanmasına yardımcı olacaktır.



Şekil 4.4. Müşteri analizi veri hazırlama programı

Veri seçimimizin ilk aşamasında 2 farklı marketteki satış hareketleri incelenerek, bu marketlerde alışveriş yapan müşterilerin satın alma eğilimleri incelenmiştir. Ürün çeşitliliğinin fazla olmasından dolayı çıkarılacak birliktelik kuralında ürünlere ait Kod_3 ürün grupları kullanılmıştır. Analiz sürecinin ilk aşamalarında ortaya çıkarılan birliktelik kuralı değerlerinin bu marketler için karşılaştırılması yapılmıştır. Daha sonrasında bu iki şubeye ait veriler tüm marketlerin ortalama verileri ile kıyaslanmıştır. Ayrı ayrı değerlendirilecek olan marketler şube 1003 ve şube 1004 olmak üzere kodlanmıştır.

4.4. R ile Apriori Algoritmasının Uygulanması

R programlama dili ile ilgili önceki bölümlerde gerekli açıklamalar yapılmıştır. R-Studio, R programlama dilinin daha rahat ve dinamik olarak kullanılması için geliştirilmiş ile bir açık kaynak kodlu tamamen ücretsiz olan bir IDE dir. Bireysel ve kurumsal kullanıcılar için oldukça zengin veri madenciliği kütüphanelerini bünyesinde hazır olarak sunmaktadır [58]. Apriori algoritmasının çalışması için birliktelik kuralı kütüphanelerin sisteme dahil edilmesi gerekmektedir.

Analiz etmek istediğimiz gerekli verileri hazırladıktan sonra sırası ile aşağıdaki r kodları ile süreç başlatılmıştır.

- Marketdata1003.txt dosyası içerisindeki satış hareketlerinin sisteme aktarılması.
  ```
  veriler = read.csv("C:\\Users\\furkan-PC\\Desktop\\marketdata1003.txt")
  ```
- Sisteme aktarılmış ilk 30 satış hareketini gösterir
  ```
  head(veriler, n=30)
  ```
- Verinin transaction olarak ayrıştırılmadan önce bölünme işlemini gerçekleştirir.
  ```
  veri1 <- split(veriler$GRUPKOD, veriler$FISID)
  ```
- Eğer birliktelik kuralı için gerekli kütüphane yüklü değilse, kütüphane'yi ekler
  ```
  if(!require(arules)) install.packages("arules")
  ```
- Verileri transaction formatına dönüştürür ve bu transaction'lar ile ilgili istatistiksel bilgileri ekrana verir.
  ```
  verit2= as(veri1,"transactions")
  summary(verit2)
  inspect(verit2)
  ```

Şekil 4.5 de frekansı en yüksek olan 5 ürün grubuna ait sayısal veriler belirtilmektedir. Bu verilere göre BİSKÜVİ ürün grubuna ait en az bir ürün 10988 farklı sepet içerisinde tercih edilmiştir.

```
most frequent items:
 NORMAL YUMURTALAR            BİSKÜVİ             PİLİÇ  ŞEKER-TATLANDIRICI      AÇIK DANA ETİ            (other)
             13193              10988               9445               9372               9180             275036
```

Şekil 4.5. Frekansı en yüksek ürünler

Şekil 4.6. da tek sepet içerisinde(tek fiş içerisinde) kaç farklı ürün grubunun geçtiğini göstermektedir. Örnek verecek olursak içerisinde sadece 1 ürün grubu bulunan sepet sayısı 8914 olurken içerisinde 10 farklı ürün grubu bulunan 1385 farklı sepet tespit edilmiştir.

```
element (itemset/transaction) length distribution:
sizes
   1    2    3    4    5    6    7    8    9   10   11   12   13   14   15   16   17   18   19   20
8914 10887 9499 7370 5412 4178 2936 2229 1721 1385  996  872  697  577  509  444  419  359  296  241
  21   22   23   24   25   26   27   28   29   30   31   32   33   34   35   36   37   38   39   40
 211  207  174  157  124  107   93   65   65   68   35   46   37   40   28   25   20   13   11   12
  41   42   43   44   45   46   47   48   51   52   54
   2    6    2    5    3    3    1    2    1    3    1
```

Şekil 4.6. Transaction'daki farklı Kod_3 grup sayıları

Şekil 4.7. de tek sepet içerisinde en az, en çok ve ortalama olarak kaç farklı ürün grubunun(Kod_3) geçmekte olduğu belirtilmektedir.

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   1.00    2.00    4.00    5.32    6.00   54.00
```

Şekil 4.7. Ortalama Kod_3 grup sayıları

Şekil 4.8. de en yüksek frekansa sahip 10 ürün grubunun toplam ürün grupları arasındaki yüzde oranı verilmektedir. 0.20 değeri % 20 değerini temsil etmektedir.



Şekil 4.8. En yüksek sayıda bulunan 10 ürün

- Apriori algoritmasının temel parametreleri olan destek(Support) ve güven(Confidence) değerleri aşağıdaki şekilde tanımlanmaktadır. Destek değerini 0.001 olarak atamamız durumunda destek değeri %0,1 oranının yukarısında tespit edilen kuralları tespit etmiş olacağız. Güven değeri 0,8 ise bize %80 ve üzeri güven değerine sahip kuralları verecektir.

kurallar = apriori(verit2, parameter=list(support=0.001, confidence=0.8))

- Tespit edilen kurallar için veri çerçevesi oluşturur. Veri çerçevesi verileri filtreleme işlemlerini kolaylaştırmaktadır.

  kurallard1 = as(kurallar, "data.frame")

- Destek ve Güven değerlerinin virgülden sonra kaç basamaklı olarak gösterileceğini belirtmektedir.

  options(digits=2)

- Şekil 4.9 Destek(Support), Güven(Confidence) ve Kaldıraç(Lift) değerleri için en az, en çok ve ortalama olarak sahip olduğu değerleri belirtmektedir.

  inspect(kurallar[1:10])

  summary(kurallar)

```
summary of quality measures:
    support           confidence          lift
 Min.   :0.001008   Min.   :0.8000   Min.   : 3.730
 1st Qu.:0.001073   1st Qu.:0.8101   1st Qu.: 5.354
 Median :0.001171   Median :0.8247   Median : 5.850
 Mean   :0.001295   Mean   :0.8307   Mean   : 5.759
 3rd Qu.:0.001382   3rd Qu.:0.8448   3rd Qu.: 6.321
 Max.   :0.010649   Max.   :1.0000   Max.   :21.975
```

Şekil 4.9. Destek, güven ve lift istatistikleri

- Verilen parametreye göre sıralama işlemi gerçekleştir. Aşağıdaki komut destek değerini göz önüne alarak büyükten-küçüğe birliktelik kurallarını dizmektedir.

  kurallar<-sort(kurallar, by="support", decreasing=TRUE)

- Daha önceki aşamalarda belirtilmiş olan koşulları sağlayan ilk 50 birliktelik kuralını oluşturmaktadır.

  inspect(kurallar[1:50])

Çizelge 4.3. de 1003 kodlu markette yapılan alışverişlere ait birliktelik kuralları verilmiştir. Birliktelik kuralları destek değerlerine göre büyükten küçüğe sıralanmıştır. Farklı Destek ve Güven değerleri için yüksek miktarda birliktelik bağlantısı tespit edilmiştir. Çizelge 4.3. de Destek = 0.001 ve Güven = 0.8 eşik değerlerini sağlayan ve en yüksek destek değerlerine sahip ilk 10 birliktelik kuralı verilmiştir.

Çizelge 4.3. deki birliktelik kurallarından en yüksek destek değerine sahip 1. Kural ve en yüksek Güven değerine sahip 9. Kural koyu renkte belirtilmiştir. Örneğin 1. Birliktelik kuralına göre Gofret ve Kek ürün gruplarına dahil olan ürünlerden her ikisinden de en az bir tane satın alan müşteriler aynı zamanla % 80,57 ihtimal ile Bisküvi ürün grubundan da en az bir ürün satın almıştır. Aynı zamanda Gofret, Kek ve Bisküvi ürün gruplarından tek sepet içerisinde en az bir kere satın alınan alışveriş sayısının toplam alışveriş sayısı içerisindeki oranı % 1,06 dır. Yani her 100 alışverişten 1,06 sın da Gofret,

Kek ve Bisküvi ürün gruplarına dahil olan ürünlerden en az bir tanesi müşteri tarafından aynı sepette tercih edilmiştir. Lift değeri ise değişkenler arasındaki kuvvetli ilişkileri göstermektedir. Kuvvetli ilişkiler için lift değerinin 1 in üzerinde olması beklenir. Aşağıdaki formül ile hesaplanmaktadır.

$$Lift\ (X\ > Y)\ =\ Güven(X\ > Y)\ /\ Destek(Y)$$

Kural 9 ise bu 10 birliktelik kuralı içerisinde en yüksek güven değerine sahiptir. Bu kurala göre Açık Dana Eti, Ketçap- Hazır Soslar, Patates-Soğan-Sarımsak ve Tereyağı ürün gruplarına dahil olan ürünlerden en az bir tane tercih eden müşteriler aynı sepette Mevsim Sebzeleri grubundan da en az bir ürünü %98,06 ihtimal ile satın almıştır. Açık Dana Eti, Ketçap- Hazır Soslar, Patates-Soğan-Sarımsak, Tereyağı ve Mevsim Sebzeleri ürün gruplarının tercih edildiği alışveriş sayısının toplam alışveriş içerisindeki payı ise %0,56 yani her 1000 alış verişin 5,6 sın da müşteriler böyle bir satın alma eğiliminde bulunmuşlardır.

Çizelge 4.3. Şube 1003 birliktelik kuralları

Kural No	Bağlantılı Ürün Grupları	Destek Oranı	Güven Oranı	Lift Oranı
1	**{GOFRET,KEK} => {BİSKÜVİ}**	**1,06%**	**80,57%**	**4,51**
2	{AÇIK DANA ETİ,OTLAR-YEŞİLLİKLER,PATATES-SOĞAN-SARIMSAK} => {MEVSİM SEBZELERİ}	0,84%	83,88%	8,46
3	{KETÇAP-HAZIR SOSLAR,PATATES-SOĞAN-SARIMSAK} => {MEVSİM SEBZELERİ}	0,81%	88,83%	8,96
4	{AÇIK DANA ETİ,PATATES-SOĞAN-SARIMSAK,TEREYAĞI} => {MEVSİM SEBZELERİ}	0,79%	86,27%	8,70
5	{KETÇAP-HAZIR SOSLAR,MEVSİM SEBZELERİ,PATATES-SOĞAN-SARIMSAK} => {AÇIK DANA ETİ}	0,70%	86,03%	5,76
6	{AÇIK DANA ETİ,KETÇAP-HAZIR SOSLAR,PATATES-SOĞAN-SARIMSAK} => {MEVSİM SEBZELERİ}	0,70%	95,99%	9,68
7	{AÇIK DANA ETİ,KETÇAP-HAZIR SOSLAR,MEVSİM SEBZELERİ} => {PATATES-SOĞAN-SARIMSAK}	0,70%	88,32%	15,54
8	{KETÇAP-HAZIR SOSLAR,PATATES-SOĞAN-SARIMSAK,TEREYAĞI} => {MEVSİM SEBZELERİ}	0,63%	96,53%	9,74
9	**{AÇIK DANA ETİ,KETÇAP-HAZIR SOSLAR,PATATES-SOĞAN-SARIMSAK,TEREYAĞI} => {MEVSİM SEBZELERİ}**	**0,57%**	**98,06%**	**9,89**
10	{AÇIK DANA ETİ,KETÇAP-HAZIR SOSLAR,MEVSİM SEBZELERİ,TEREYAĞI} => {PATATES-SOĞAN-SARIMSAK}	0,57%	93,63%	16,48

Çizelge 4.4. Şube 1003'de bisküvi ürün grubuna ait birliktelik kuralları

Kural No	Bağlantılı Ürün Grupları	Destek Oranı	Güven Oranı	Lift Oranı
1	{GOFRET,KEK} => {BİSKÜVİ}	1,06%	80,60%	4,51
2	{KEK,MAKARNA,PASTA MALZEMELERİ} => {BİSKÜVİ}	0,39%	80,90%	4,53
3	{KEK,KREM ÇİKOLATA-EZME,MAKARNA} => {BİSKÜVİ}	0,36%	81,00%	4,54
4	{AYÇİÇEK YAĞLARI,KEK,MAKARNA} => {BİSKÜVİ}	0,36%	81,40%	4,56
5	{GOFRET,KEK,NORMAL YUMURTALAR} => {BİSKÜVİ}	0,35%	83,20%	4,66
6	{KEK,MAKARNA,ŞAMPUANLAR} => {BİSKÜVİ}	0,35%	83,50%	4,68
7	{KEK,MAKARNA,NORMAL PEÇETELER,NORMAL YUMURTALAR} => {BİSKÜVİ}	0,34%	80,70%	4,52
8	{KEK,MAKARNA,TEREYAĞI} => {BİSKÜVİ}	0,34%	80,50%	4,51
9	{ÇİKOLATA,KEK,MEYVE SUYU} => {BİSKÜVİ}	0,33%	80,10%	4,48
10	**{GOFRET,KEK,MAKARNA} => {BİSKÜVİ}**	**0,31%**	**88,40%**	**4,95**

- Eğer belirli bir ürün grubuna ait birliktelik bağlantısı çıkarılmak istenirse, aşağıdaki R kodu ile bu işlem gerçekleştirilebilmektedir.

  inspect(subset(kurallar, subset = rhs %pin% "BİSKÜVİ"))

Genel Birliktelik kurallarının yanı sıra firma belirli bir ürün grubunun müşteriler tarafında hangi ürün grupları ile birlikte tercih edildiğini öğrenmek isteyebilir. Örneğin Çizelge 4.4.'de Bisküvi grubuna ait birliktelik kuralları verilmiştir. Bu çizelgede sadece Bisküvi grubu ile beraber tercih edilen ürün grupları verilmiştir. En yüksek Güven değerine sahip olan 10.kuralı açıklayacak olursak tek sepette Gofret, Kek, Makarna gruplarından her birine ait en az bir ürün satın alan müşteri %88,4 oranla Bisküvi grubundan da en az bir ürün satın almıştır. Bununla birlikte tek sepette Gofret, Kek, Makarna ve Bisküvi ürün gruplarının her birinden en az bir adet ürünün tercih edildiği alışveriş sayısının toplam alışveriş içerisindeki oranı % 0,31 dir. Belirli bir ürün grubu üzerinde elde edilen bu birliktelik kuralları satıcı firma açısından, özellikle kampanya ve promosyon oluşturma süreçlerinde önemli avantajlar sağlayacaktır.

Çizelge 4.5. de 1004 kodlu markete ait birliktelik kuralları verilmiştir.1004 kodlu market ve 1003 ayrı lokasyonlar'da olup her iki market için çıkarılacak birliktelik kuralları,

iki market arasındaki müşteri profili arasındaki farklılıkların belirlenmesinde önemli bilgiler sunmuştur. Çizelge 4.5. 'de birliktelik kuralları destek değerlerine göre büyükten küçüğe olarak sıralanmıştır ve en yüksek destek değerine sahip ilk 10 kural çizelgede verilmiştir. 1. Kuralı incelediğimiz zaman Gofret ve Kek ürün grubundan en az bir ürün satın alan müşteriler aynı zamanda Bisküvi ürün grubundan da en az bir ürünü % 81,50 ihtimal ile satın alınmıştır. Aynı zamanda Gofret, Kek, Bisküvi ürün gruplardan en az bir ürünün tercih edildiği alışveriş sayısının bütün alışveriş içerisindeki oranı %1,71'dir.

Çizelge 4.5. Şube 1004'e ait birliktelik kuralları

Kural No	Bağlantılı Ürün Grupları	Destek Oranı	Güven Oranı	Lift Oranı
1	**{GOFRET,KEK} => {BİSKÜVİ}**	**1,71%**	**81,50%**	**4,53**
2	{ÇAY,PİRİNÇ,SİYAH ZEYTİNLER} => {ŞEKER-TATLANDIRICI}	1,49%	80,40%	4,27
3	{ÇAY,MAKARNA,NORMAL YUMURTALAR,PİRİNÇ} => {ŞEKER-TATLANDIRICI}	1,48%	80,70%	4,28
4	{ÇAY,MAKARNA,NORMAL PEÇETELER,PİRİNÇ} => {ŞEKER-TATLANDIRICI}	1,43%	81,10%	4,31
5	{ÇAY,MAKARNA,SIVI DETERJAN} => {ŞEKER-TATLANDIRICI}	1,37%	80,80%	4,29
6	{ÇAY,KREM ÇİKOLATA-EZME,MAKARNA} => {ŞEKER-TATLANDIRICI}	1,35%	80,80%	4,29
7	{ÇAY,PİRİNÇ,SALÇA} => {ŞEKER-TATLANDIRICI}	1,34%	82,50%	4,38
8	{AYÇİÇEK YAĞLARI,ÇAY,MAKARNA,NORMAL YUMURTALAR} => {ŞEKER-TATLANDIRICI}	1,25%	80,40%	4,27
9	{ÇAY,PİRİNÇ,TEREYAĞI} => {ŞEKER-TATLANDIRICI}	1,23%	83,00%	4,40
10	{ÇAY,SALÇA,SİYAH ZEYTİNLER} => {ŞEKER-TATLANDIRICI}	1,22%	80,20%	4,26

Çizelge 4.6. da 1004 kodlu marketin satın alma verilerine göre müşterilerin bisküvi ürün grubundan satın almış oldukları ürünleri başka hangi ürün gruplarındaki ürünler ile birlikte tercih ettiklerine dair bazı birliktelik kuralları verilmiştir. Bisküvi ürün grubunun Gofret, Kek, Makarna ile tercih edilme oranı % 86,30 iken, Kek ve Sakız ürün grubu tercih edildikten sonra birlikte tercih edilme oranı 83,20 olarak tespit edilmiştir.

Çizelge 4.6. Şube 1004'de bisküvi ürün grubuna ait birliktelik kuralları

Kural No	Bağlantılı Ürün Grupları	Destek Oranı	Güven Oranı	Lift Oranı
1	{GOFRET,KEK} => {BİSKÜVİ}	1,71%	81,50%	4,53
2	{KEK,MAKARNA,NORMAL PEÇETELER} => {BİSKÜVİ}	1,02%	80,70%	4,49
3	{KEK,SAKIZ} => {BİSKÜVİ}	**0,95%**	**83,20%**	**4,62**
4	{ÇAY,KEK,NORMAL PEÇETELER} => {BİSKÜVİ}	0,82%	81,50%	4,53
5	{GOFRET,KEK,NORMAL YUMURTALAR} => {BİSKÜVİ}	0,78%	82,80%	4,60
6	{KEK,PİLİÇ ,ŞEKER-TATLANDIRICI} => {BİSKÜVİ}	0,77%	80,70%	4,49
7	{KEK,MAKARNA,PASTA MALZEMELERİ} => {BİSKÜVİ}	0,75%	82,00%	4,56
8	**{GOFRET,KEK,MAKARNA} => {BİSKÜVİ}**	**0,71%**	**86,30%**	**4,80**
9	{KEK,SİYAH ZEYTİNLER,ŞEKER-TATLANDIRICI} => {BİSKÜVİ}	0,70%	81,00%	4,50

4.5 Birliktelik Kurallarına Ait Müşteri Bilgilerinin Tespiti

Daha önceki bölümlerde iki farklı markete ait satış verileri üzerinden birbiri ile bağlantılı grupların tespiti yapılmıştır. Fakat şu an bu veriler anonim veriler olarak bulunmaktadır. Yani bu veriler promosyon ve kampanyalar oluşturmak için değerli bilgilerdir. Fakat müşterilerin ürünleri tercih etme profilini daha ayrıntılı bir şekilde ortaya çıkarmak için yeterli değildir. Bu aşamada müşterin kişisel bilgileri ile bu kurallar arasındaki bağlantıyı ortaya çıkarıp kampanya ve promosyonların hazırlanmasında nokta atışı yaparak hedef kitlenin belirlenmesinde daha sağlıklı sonuçlar elde edilmesi amaçlanmıştır. İlk aşamada 1003 ve 1004 kodlu marketlere ait birliktelik kuralları göz önüne alınarak bu marketlerden 1. Yıl içerisinde alışveriş yapmış müşterilerin özellikleri çıkarılacaktır. Çizelge 4.7. de görüldüğü üzere iki markette de en güçlü ortak birliktelik kuralı verilmiştir. Seçilen kuralın seçilme sebebi 100 alışveriş içerisinden en az 1 alışveriş içerisinde gerçekleşmesidir. Bu sayede elde edilen sonuçların yönlendirici ve belirleyici olması hedeflenmiştir. Bu bölümde ortaya çıkarılacak veriler {GOFRET,KEK} => {BİSKÜVİ} birliktelik kuralını oluşturan müşterilere ait olacaktır. Fakat bu birliktelik kuralından oluşturulacak kampanya sadece bu müşteri kitlesine yönelik olamayacaktır. Bu çalışmanın asıl amaçlarından biriside benzer alışverişleri gerçekleştiren müşteriler için aynı kampanyaları sunmak yani gofret ve kek ürün gruplarına dahil olan bir ürün satın aldıktan sonra bisküvi ürün grubuna ait bir ürün satın almayan müşteriler için aynı kampanyayı sunmak. Buradaki temel amaç benzer alış veriş eğilimin gösteren ve benzer kişisel özelliklere sahip müşterilere benzer kampanyalar sunmaktır.

Çizelge 4.7. Şube 1003 ve 1004 ‘e ait en yüksek destek değere sahip birliktelik kuralları

Market No	*Bağlantılı Ürün Grupları*	*Destek(Support) Oranı*	*Güven(Confidence) Oranı*	*Lift Oranı*
1003	{GOFRET,KEK} => {BİSKÜVİ}	1,06%	80,60%	4,51
1004	{GOFRET,KEK} => {BİSKÜVİ}	1,71%	81,50%	4,53

Elde edilen birliktelik kurallarına ait müşteri profillerinin çıkarılması ve müşteri özelliklerinin daha kolay analiz edilebilmesi için Mssql, Devexpress Dasboard, C# teknolojileri kullanılarak bir uygulama yapılmıştır. Uygulama aşağıdaki modüller den oluşmaktadır. Elde edilen sonuçların daha hızlı ve kolay analiz edilebilmesi için hazırlanan program önemli avantajlar sağlamaktadır.

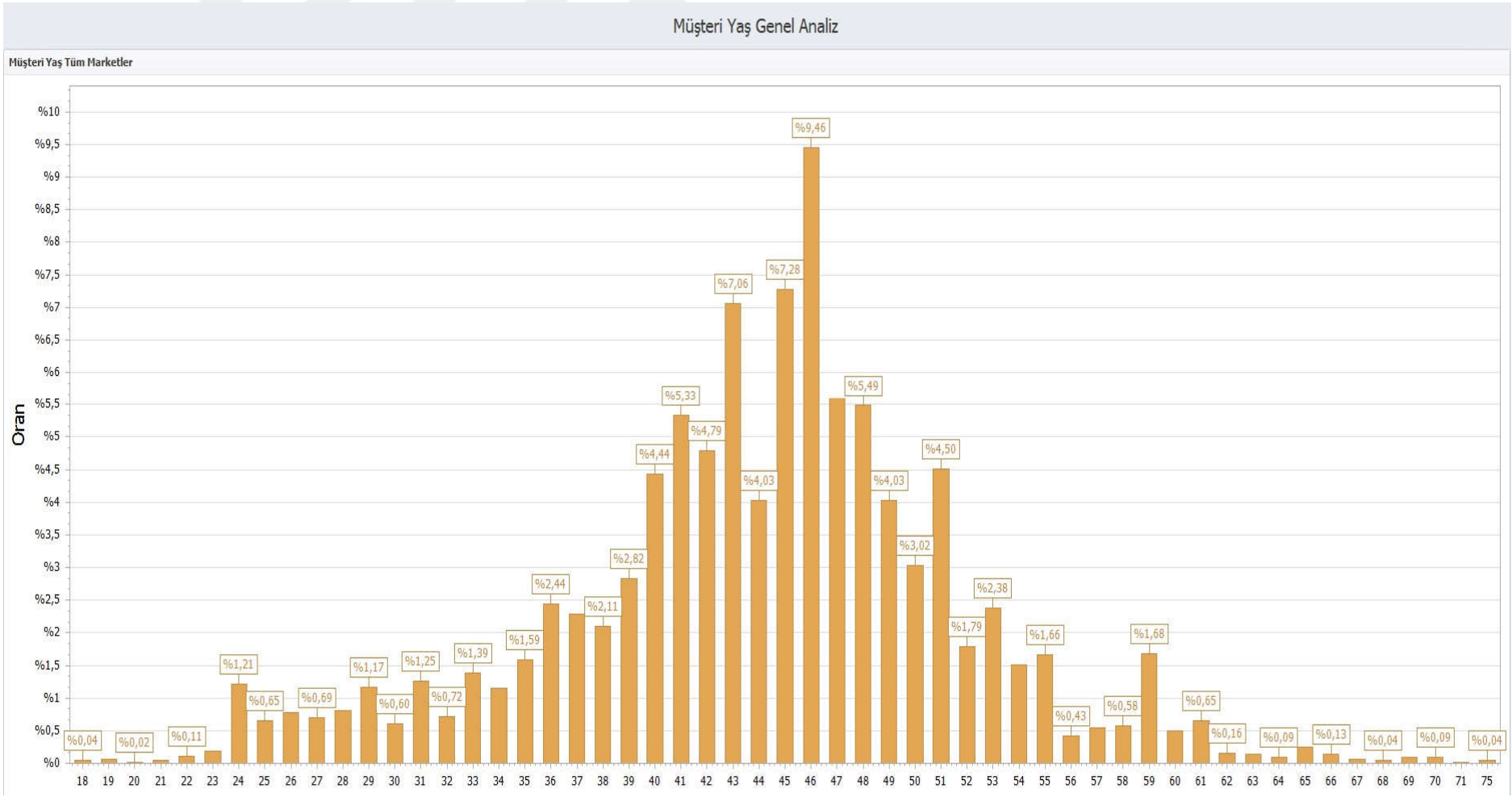
- Müşteri Öğrenim Bilgileri Modülü
- Müşteri Cinsiyet Bilgileri Modülü
- Müşteri Yaş Bilgileri Modülü
- Müşteri Adres Bilgileri Modülü
- Müşteri Tarih Bilgileri Modülü
- Müşteri Alışveriş Bilgileri Modülü
- Müşteri Meslek Bilgileri Modülü
- Müşteri Medeni Hâl Bilgileri Modülü
- Müşteri Sepet Tutar Analiz Modülü

Şekil 4.10. da belirtilen marketlerdeki müşterilerin 1003 ve 1004 nolu marketlerden Gofret, Kek ve Bisküvi ürün gruplarındaki ürünlerden en az bir tane satın alan müşterilere ait yaş profilleri verilmiştir. Şube 1003 deki müşterilerden bu ürün gruplarını tercih eden müşterilerin %17,61 ‘i 48 yaşındadır. % 12,58 ‘i 43 yaşında %8,81 i ise 45 yaşındadır. Yaş aralıklarına göre gruplandırıldığında 40-50 yaşları arasındaki müşteri gruplarının oranının %66,81 olduğu görülmektedir. Aynı zamanda bu yaş aralığı açık ara en büyük yüzdelik orana sahip olduğu için hedef kitlenin belirlenmesinde önemli bir parametredir. Şube 1004 de müşterilerin % 14,52 si 51 yaşında ,%13,20 si 47 yaşında, %10,23 ‘ü 46 yaşındadır. Grupsal olarak baktığımızda ise 40-50 yaşları aşasındaki müşteriler toplam müşteriler içerisindeki %39,27’lik oranı oluşturmaktadır. Ayrıca bütün şubelerin kümülatif değerleri üzerinden ortalama müşteri yaş profili Şekil 4.11 de verilmiştir. Şekil 4.11. deki grafiğe göre ;

Gofret, Kek ve Bisküvi ürün gruplarından en az bir ürünü sepetinde bulunduran müşterilerin % 9,46 ‘sı 46 yaşındadır. % 7,28’i 45 yaşındadır. %7,06’sı 43 yaşındadır. 40-50 yaşları arasındaki müşteri grubu ise toplam müşterilerin %57,61’ini oluşturmaktadır.



Şekil 4.10. Müşteri yaş dağılım karşılaştırması



Şekil 4.11. Belirtilen ürün gruplarına ait yaş aralığı(Tüm Şubeler)

Şekil 4.12. de 1003 ve 1004'nolu marketlerde {GOFRET, KEK} => {BİSKÜVİ} birliktelik kuralını sağlayan müşterilere ait Eğitim Bilgileri verilmiştir. Ayrıca tüm marketlere ait eğitim bilgileri grafiği ise Şekil 4.13. 'de verilmiştir.

Alışverişlerinde belirtilen ürün gruplarını tercih etmiş olan müşterilerin eğitim durumları aşağıdaki çizelgelerde verilmiştir.

Çizelge 4.8. Şube 1003 müşteri eğitim oranı

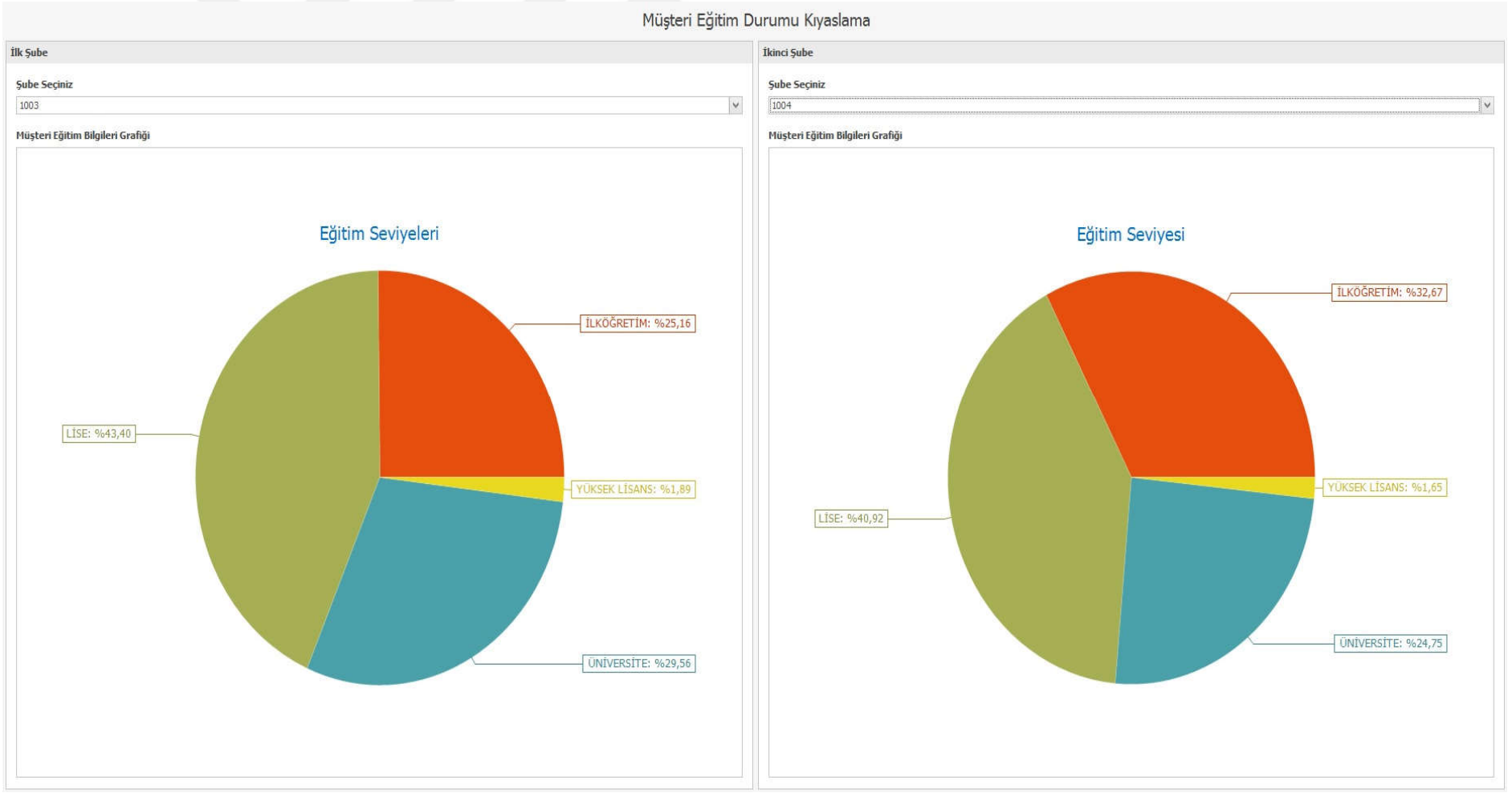
Eğitim Durumu	Yüzde Oran
Lise	% 43,40
Üniversite	% 29,56
İlköğretim	% 25,16
Yüksek Lisans	% 1,89

Çizelge 4.9. Şube 1004 müşteri eğitim oranı

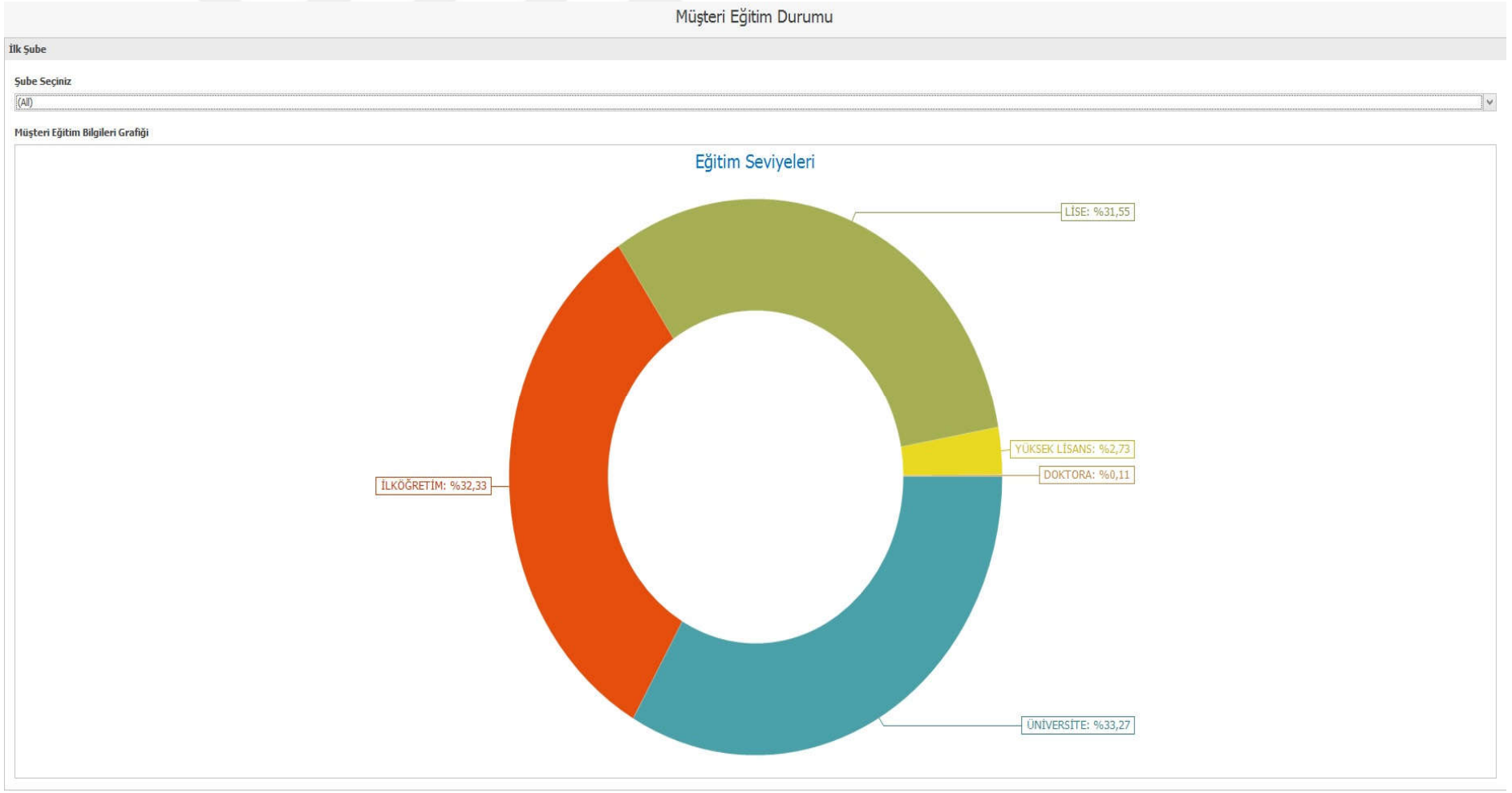
Eğitim Durumu	Yüzde Oran
Lise	% 40,92
İlköğretim	% 32,67
Üniversite	% 24,75
Yüksek Lisans	% 1,65

Çizelge 4.10. Bütün şubeler eğitim oranları

Eğitim Durumu	Yüzde Oran
Üniversite	% 33,27
İlköğretim	% 32,33
Lise	% 31,55
Yüksek Lisans	% 2,73
Doktora	% 0,11



Şekil 4.12. Farklı marketlerdeki eğitim durumu karşılaştırması



Şekil 4.13. Tüm şubeler eğitim seviyesi grafiği

Şekil 4.14 de 1003 ve 1004'nolu marketlerde Gofret, Kek ve Bisküvi ürün gruplarındaki ürünlerden en az bir tane satın alan müşterilere ait en yüksek yüzdelik orana sahip olan ilk 20 meslek grubu verilmiştir. Şekil 4.15. de aynı birliktelik kurallı için bütün şubeleri kapsayan meslek bilgileri verilmiştir. Ayrıca aşağıdaki çizelgelerde ise en yüksek orana sahip 6 meslek grubuna ait bilgiler verilmiştir.

Çizelge 4.11. Şube 1003 meslek grupları

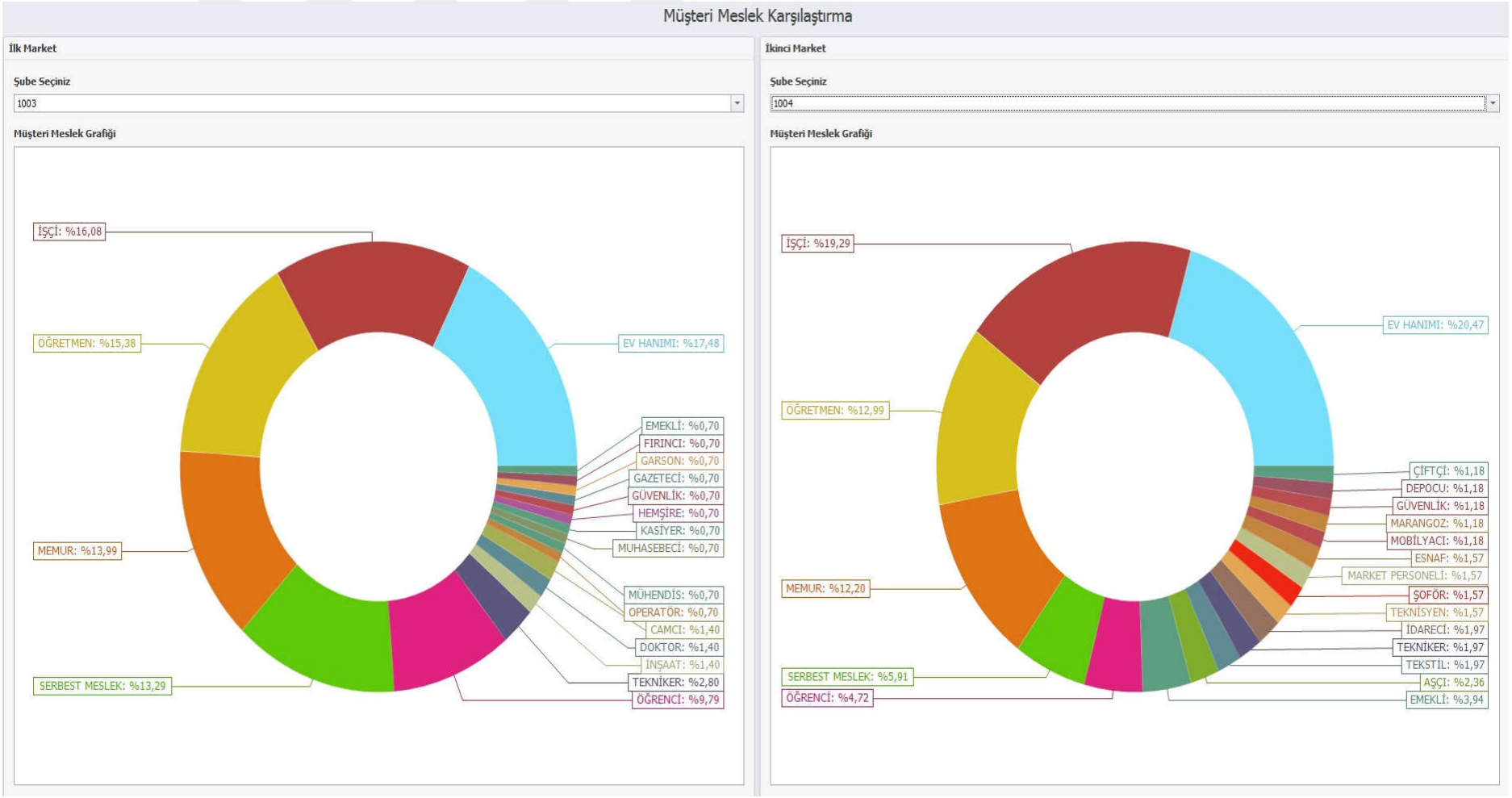
Meslekler Grupları	Yüzde Oran
Ev Hanımı	% 17,48
İşçi	% 16,08
Öğretmen	% 15,38
Memur	% 13,99
Serbest Meslek	% 13,29
Öğrenci	% 9,79

Çizelge 4.12. Şube 1004 meslek grupları

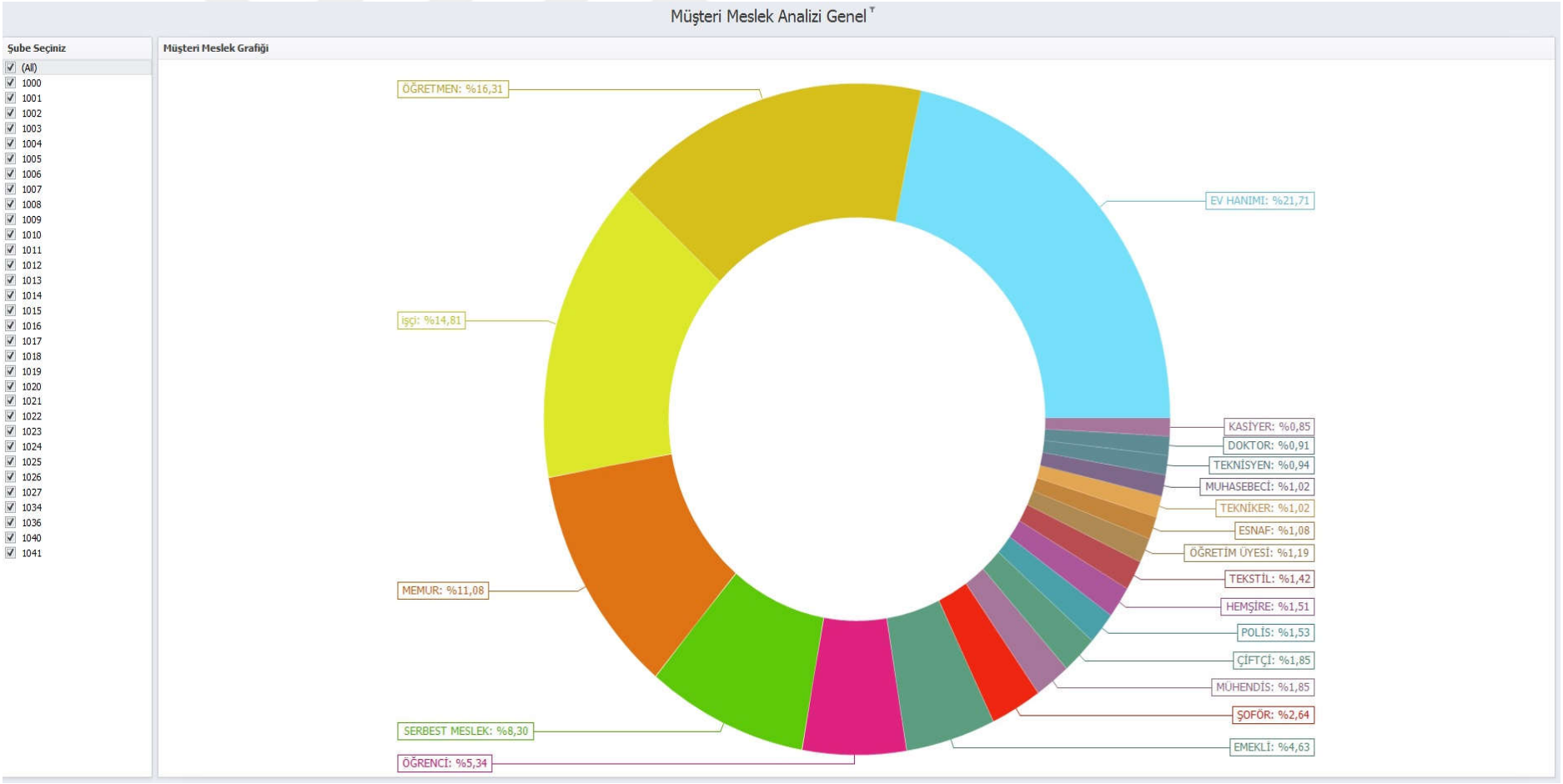
Meslek Grupları	Yüzde Oran
Ev Hanımı	% 20,47
İşçi	% 19,29
Öğretmen	% 12,99
Memur	% 12,20
Serbest Meslek	% 5,91
Öğrenci	% 4,72

Çizelge 4.13. Bütün şubeler meslek grupları

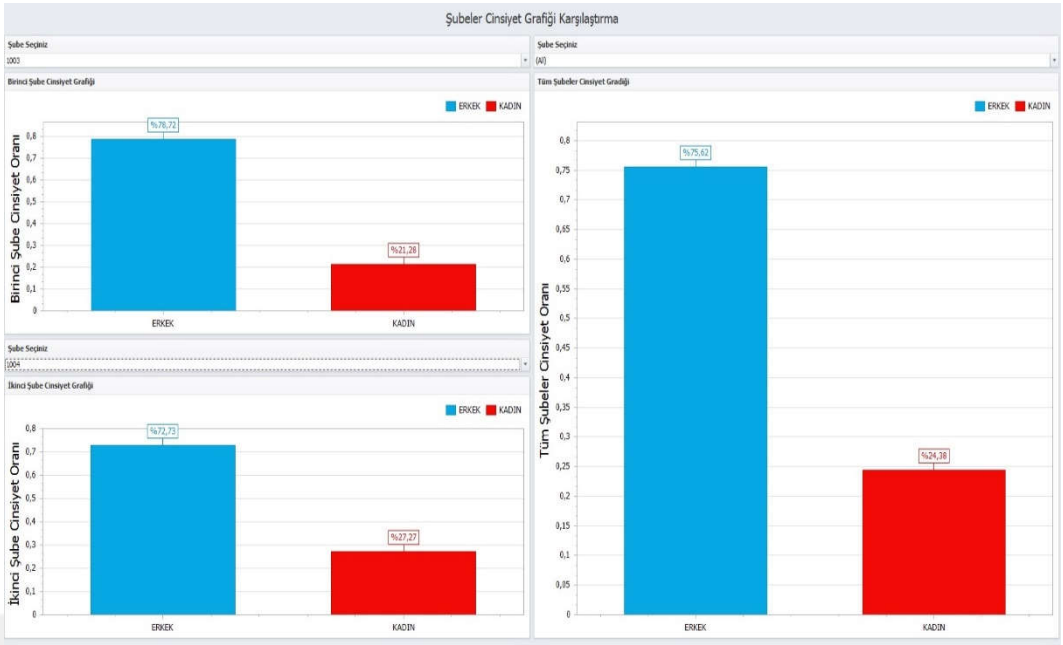
Meslek Grupları	Yüzde Oran
Ev Hanımı	% 21,71
Öğretmen	% 16,31
İşçi	% 14,81
Memur	% 11,08
Serbest Meslek	% 8,30
Öğrenci	% 5,34



Şekil 4.14. Farklı şubeler meslek oranları karşılaştırma

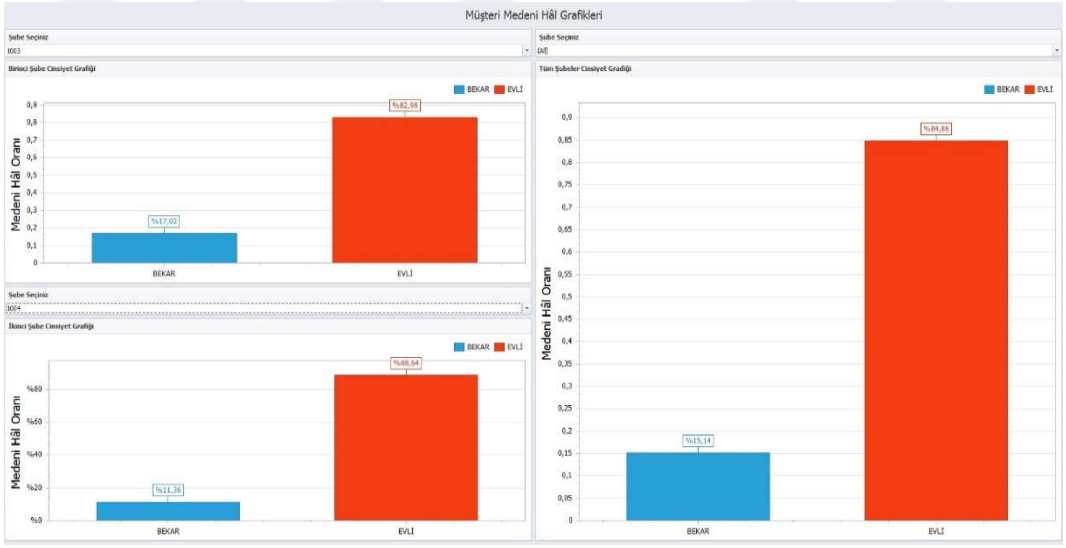


Şekil 4.15. Tüm şubelere ait meslek grupları grafiği



Şekil 4.16. Farklı şubeler müşteri cinsiyet oranları karşılaştırma

Şekil 4.16.'da 1003 nolu şubede gofret, kek ve bisküvi ürün gruplarındaki ürünlerden en az bir tane satın alan müşterilerin % 78,72 si Erkek ve % 21,28 'inin kadın olduğu tespit edilmiştir. 1004'nolu şubede ise % 72,73 Erkek ve % 27,27 Kadın oranları elde edilmiştir. Bütün marketlerin bilgilerini dahil ettiğimizde % 75,62 Erkek ve % 24,38 kadın oranlarına ulaşmaktayız.



Şekil 4.17. Farklı şubelerdeki müşterilerin medeni durumları

Şekil 4.17.'de 1003 nolu şubede gofret, kek ve bisküvi ürün gruplarındaki ürünlerden en az bir tane satın alan müşterilerin % 82,98'i Evli ve % 17,02'sinin Bekâr olduğunu söylenebilir. Şube 1004'de ise müşterilerin % 88,64'ü evli ve % 11,36'sı bekârdır. Tüm market bilgilerini dahil ettiğimizde müşterilerin % 84,86'sı Evli ve % 15,14'ü Bekar olarak tespit edilmiştir.

Belirtilen ürün gruplarının müşteriler tarafından tercih edilme durumlarını zaman açısından incelediğimizde aylık olarak şekil 4.18 de, günlük olarak şekil 4.19 ve saat olarak da şekil 4.20 gösterilen grafikler elde edilmiştir.

Şekil 4.18 de görüldüğü üzere her iki şubenin de verileri aylık olarak incelendiğinde 1003 şubesinde Gofret, Kek ve Bisküvi ürün gruplarındaki ürünlerden en az bir tane satın alan müşteriler yaptıkları alışverişlerin %13.19 Kasım ayında, %12.03'ü Eylül ayında, %11.88'ininde Nisan ayında yapıldığı görülmüştür. 1004 Şubesindeki ilk 3 ayı inceleyecek olursak bu oranlar %12.64 Nisan, %11.60 Şubat ve %11.31 Ocak ayları olarak belirlenmiştir. Bütün şubeler analiz edildiğinde en yüksek orana sahip ilk 3 ay sırası ile Nisan %10.96, Kasım %9.83 ve Şubat % 9.53 olarak sıralanmaktadır

Şekil 4.19'da görüldüğü üzere haftanın günleri açısından incelendiğinde 1003 şubesinde bu ürün gruplarının Pazar günü tercih edilme oranı %19.57, Pazartesi tercih edilme oranı %15.07 ve Cuma günü tercih edilme oranı %14,20 olarak tespit edilmiştir. 1004 şubesindeki en yüksek orana sahip 3 günü incelediğimiz zaman ise Pazar günü %19.57, Pazartesi günü 15.07 ve Cuma günü %14.20 oranında belirlenen ürün grupları tercih edilmiştir.

Şekil 4.20'de belirtilen ürün gruplarının gün içerisinde hangi saat dilimlerinde ne kadar tercih edildiğine dair oranlar verilmiştir. 1003 ve 1004 nolu markette ait verilerin bir kısmı büyükten küçüğe sıralı olarak aşağıdaki çizelgelerde verilmiştir;

Çizelge 4.14. Şube 1003 saat aralıklarına göre tercih oranları

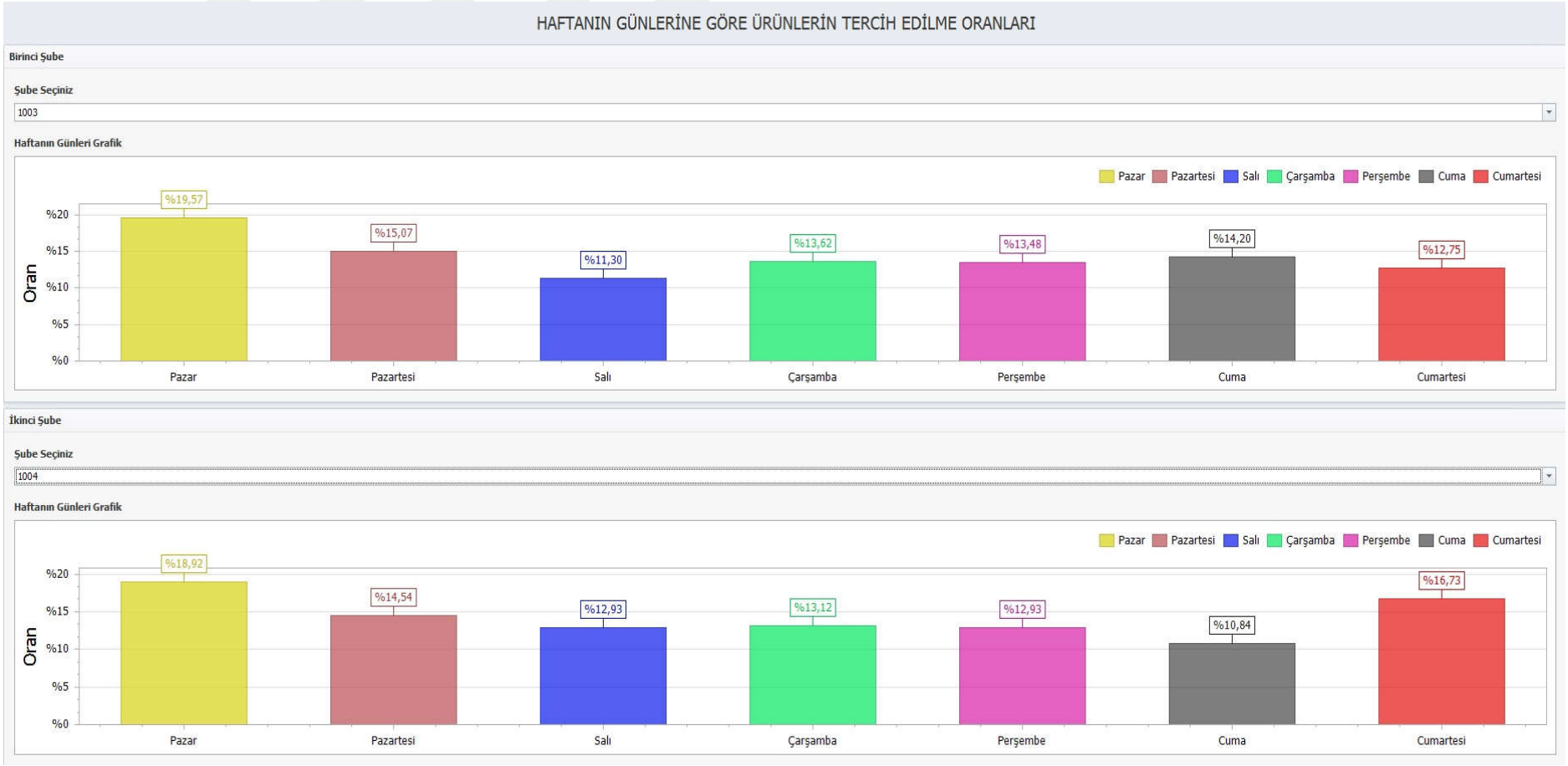
Saat aralıkları	Yüzde Oran
19.00-20.00	%13.04
20.00-21.00	%11.30
17.00-18.00	%11.01
18.00-19.00	%9.71
15.00-16.00	%8.84

Çizelge 4.15. Şube 1004 saat aralıklarına göre tercih oranları

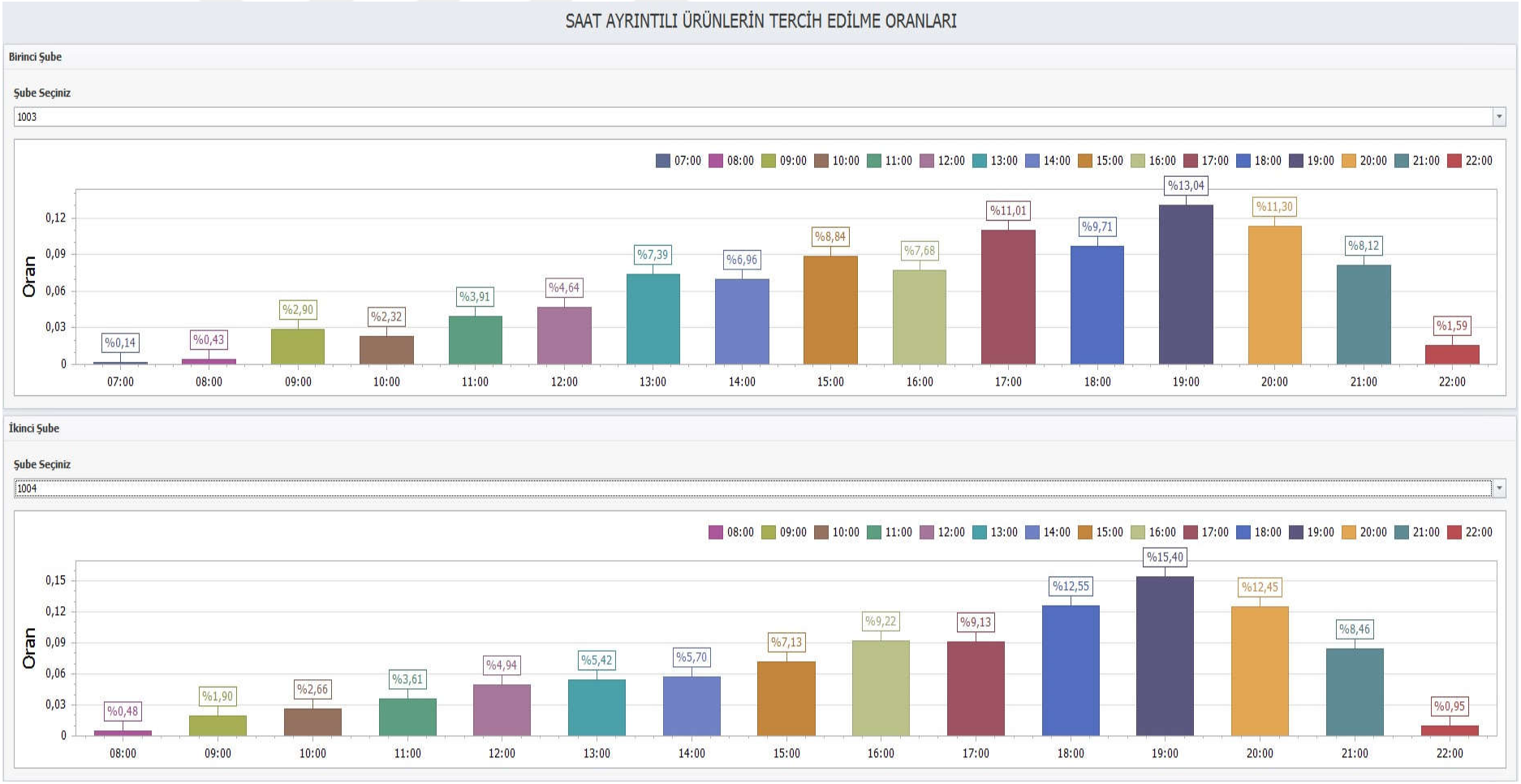
Saat aralıkları	Yüzde Oran
19.00-20.00	%15.40
18.00-19.00	%12.55
20.00-21.00	%12.45
16.00-17.00	%9.22
17.00-18.00	%9.13



Şekil 4.18. Farklı şubelerde belirtilen ürün grubunun tercih edildiği ayların karşılaştırılması



Şekil 4.19. Farklı şubelerde belirtilen ürün grubunun tercih edildiği haftanın günlerine göre karşılaştırılması



Şekil 4.20. Farklı şubelerde belirtilen ürün grubunun tercih edildiği saat aralıklarına göre karşılaştırılması

Şekil 4.21. de Gofret, Kek ve Bisküvi ürün gruplarındaki ürünlerden en az bir tane satın alan müşterilerin sepet tutarlarına ait veriler grafik üzerinde belirtilmiştir. Bu bilgiler bize belirtilen ürün gruplarını tercih eden müşterilerin bu alışverişi ne kadarlık bir sepet tutarı içerisinde gerçekleştirdiğini gösterecektir. Şube 1003, şube 1004 ve bütün şubelere ait sepet tutarlarına ait bilgiler aşağıdaki çizelgelerde büyükten küçüğe sıralı olacak şekilde en yüksek orana sahip beş grup olarak verilmiştir.

Çizelge 4.16. Şube 1003 sepet tutar oranları

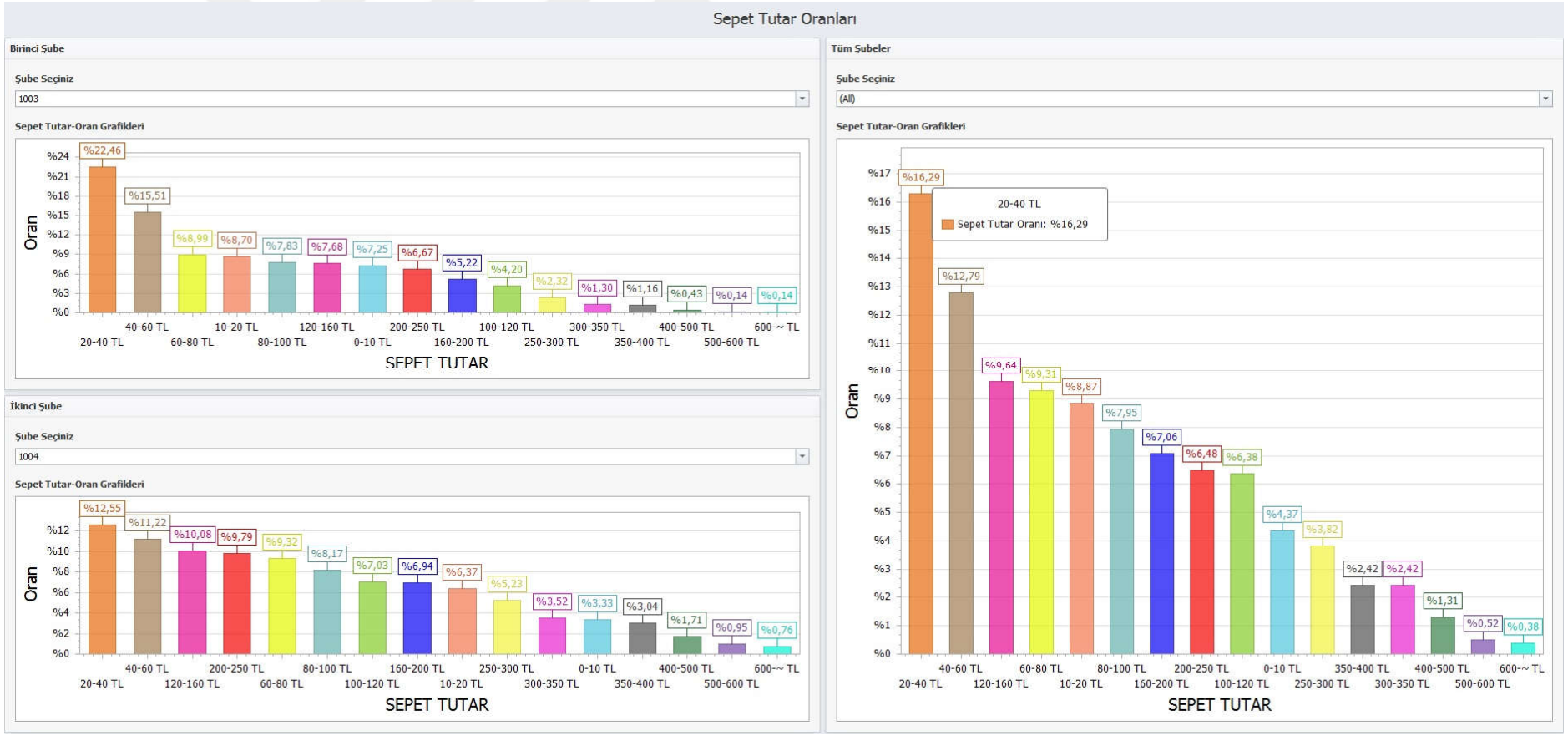
Sepet Tutar Aralığı	Yüzde Oran
20 TL - 40 TL	%22,46
40 TL - 60 TL	%15,51
60 TL - 80 TL	%8,99
10 TL - 20 TL	%8,70
80 TL - 100 TL	%7,83

Çizelge 4.17. Şube 1004 sepet tutar oranları

Sepet Tutar Aralığı	Yüzde Oran
20 TL - 40 TL	%12,55
40 TL - 60 TL	%15,51
120 TL - 160 TL	%10,08
200 TL - 250 TL	%9,79
60 TL - 80 TL	%9,32

Çizelge 4.18. Tüm şubeler sepet tutar oranları

Sepet Tutar Aralığı	Yüzde Oran
20 TL – 40 TL	%16,29
40 TL – 60 TL	%12,79
120 TL – 160 TL	%9,64
60 TL – 80 TL	%9,31
10 TL – 20 TL	%8,87



Şekil 4.21. Farklı şubeler sepet tutar oranları karşılaştırma

Şekil 4.22. Müşterilerin yapmış oldukları toplam alışveriş tutarı içerisindeki Gofret, Kek ve Bisküvi ürün gruplarına ait ürünlerin tutarsal olarak oranını vermektedir. Şube 1003 için açıklayacak olursak;

Aynı sepet içerisindeki satın alınan Gofret, Kek ve Bisküvi ürün gruplarına dâhil olan ürünlerin toplam tutarının toplam sepet tutarına oranı %0 ile %5 arasında ise bu alışverişlerin sayısının toplam alışveriş sayısına oranı % 13,48'dir. Bir başka örnek ile açıklayacak olursak müşterinin yapmış olduğu alışveriş de Gofret, Kek ve Bisküvi grubuna ait ürünlerin toplam tutarının, toplam alışverişe oranı %5 ile %8 arasında ise bu belirtilen şartları sağlayan alışverişlerin sayısı Gofret, Kek ve Bisküvi ürün gruplarına dâhil ürünlerden en az bir tane satın alınan toplam alıverişlerin % 14,64 üne denk gelmektedir. Şube 1003, Şube 1004 ve Bütün marketlerin kümülatif değerlerine göre hesaplanan en yüksek orana sahip 5 gruba ait veriler aşağıdaki çizelgelerde büyükten küçüğe göre sıralanmıştır.

Çizelge 4.19. Şube 1003 belirlenen ürünlerin sepet içerisindeki oranı

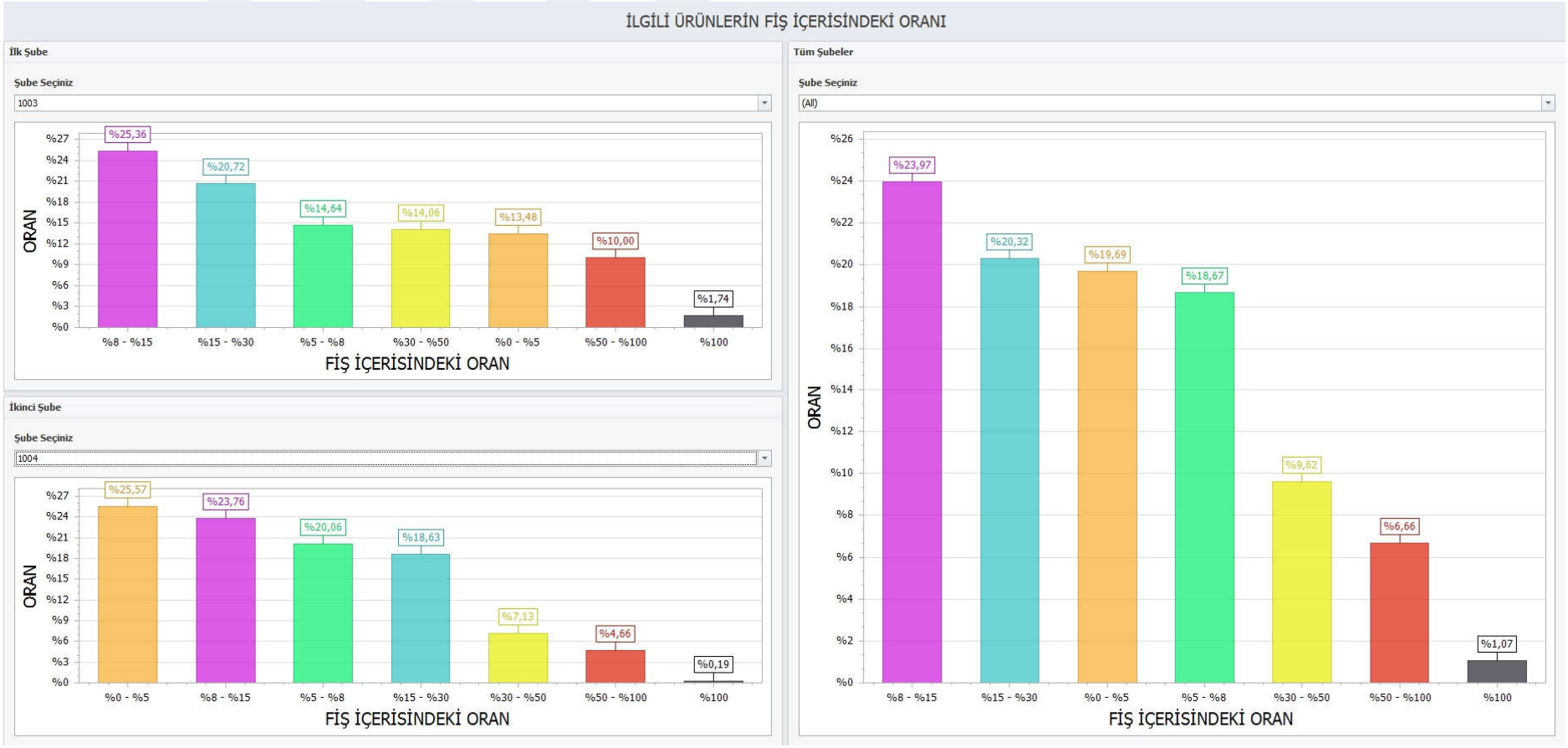
(Ürün Grubu Tutar/ Sepet Toplam Tutar) Aralığı	Toplam Sepet İçerisindeki Oran
%8-%15	%25,36
%15-%30	%20,72
%5-%8	%14,64
%30-%50	%14,06
%0-%5	%13,48

Çizelge 4.20. Şube 1004 belirlenen ürünlerin sepet içerisindeki oranı

(Ürün Grubu Tutar/ Sepet Toplam Tutar) Aralığı	Toplam Sepet İçerisindeki Oran
%0-%5	%25,27
%8-%15	%23,76
%5-%8	%20,06
%15-%30	%18,63
%30-%50	%7,13

Çizelge 4.21. Tüm şubeler belirlenen ürünlerin sepet içerisindeki oranı

(Ürün Grubu Tutar/ Sepet Toplam Tutar) Aralığı	Toplam Sepet İçerisindeki Oran
%8-%15	%23,97
%15-%30	%20,32
%0-%5	%19,69
%5-%8	%18,67
%30-%50	%9,62



Şekil 4.22. Belirlenen ürünlerin sepet içerisindeki oranlarının karşılaştırılması

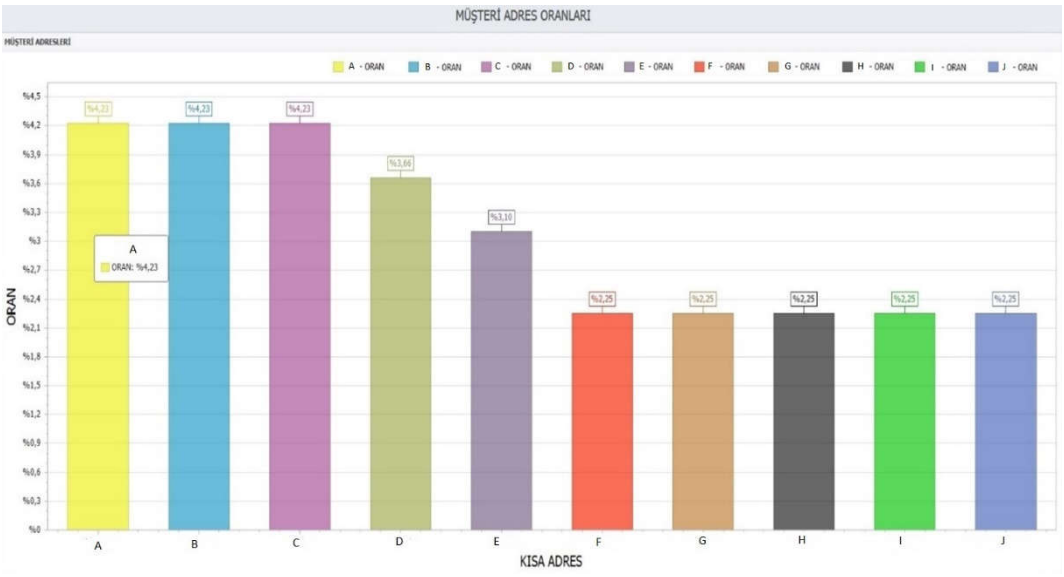
Şekil 4.23 ve Şekil 4.24 de sırası işe Şube 1003 ve Şube 1004' de belirtilen ürün gruplarına ait ürünleri birlikte satın almış olan müşterilerin adres bilgilerine ait grafik verilmiştir. Bu grafiklerde en yüksek orana sahip olan 10 adres bulunmaktadır. Bu adres bilgileri müşterilerin özel bilgileri olduğu için filtrelenerek kısıtlı olarak sunulmuştur. Çizelge 4.22'de şube 1003 ve şube 1004' e ait bilgiler verilmiştir.

Çizelge 4.22. Şube 1003 müşteri adres bilgileri

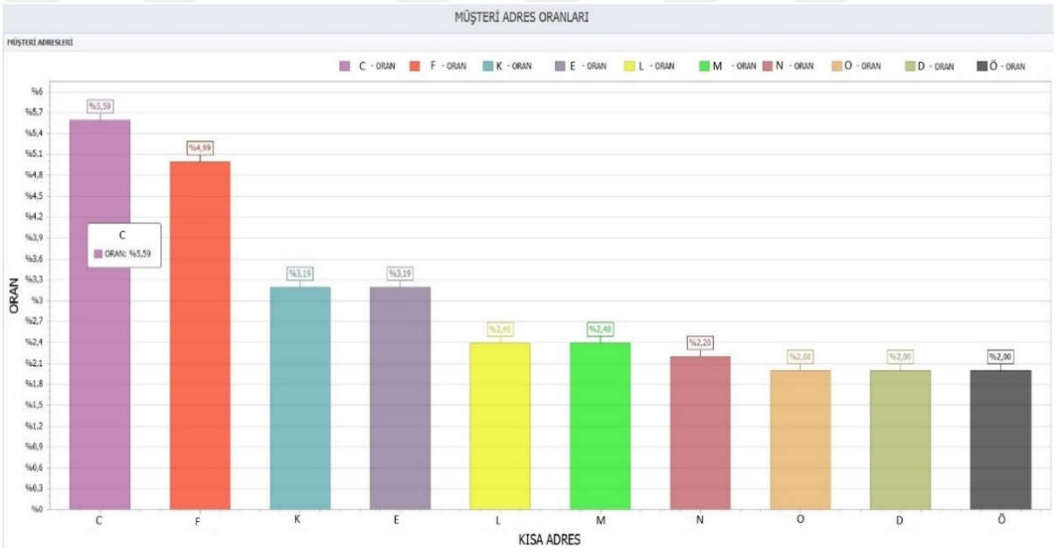
Kısa Adres	Oran
A	%4,23
B	%4,23
C	%4,23
D	%3,66
E	%3,10
F	%2,25
G	%2,25
H	%2,25
I	%2,25
J	%2,25

Çizelge 4.23. Şube 1004 müşteri adres bilgileri

Kısa Adres	Yüzde Oran
C	%5,59
F	%4,99
K	%3,19
E	%3,19
L	%2,40
M	%2,40
N	%2,20
O	%2,00
D	%2,00
Ö	%2,00



Şekil 4.23. Şube 1003 adres oranları



Şekil 4.24. Şube 1004 adres oranları

Çizelge 4.24 ’de ise market ayrıntılı olarak 1. ve 2. senelerinde gofret, kek ve bisküvi gruplarına ait en az bir ürün satın alan müşterilerin sepetlerinde bu ürünlerin ne kadar bir tutar oluşturduğu bilgisi verilmiştir. Bu sayede bu kurala yönelik uygulanan her hangi bir kampanya işleminin hangi şubelerde daha başarılı olacağına yönelik bir tahmin yapabilme imkânı oluşturulmuş olacaktır. Analiz edilen {GOFRET,KEK} => {BİSKÜVİ} birliktelik kuralının diğer şubelerdeki destek ve güven değerlerini gösteren grafik Şekil 4.25’de verilmiştir.

Çizelge 4.24 Şubelerdeki belirtilen ürün gruplarının sepetteki ortalama tutarları

ŞUBE	1. YIL	2. YIL	DEĞİŞİM
1000	11,68	15,40	32%
1001	9,15	10,27	12%
1002	10,95	12,23	12%
1003	9,57	11,61	21%
1004	10,26	12,10	18%
1005	9,52	11,79	24%
1006	9,11	11,55	27%
1007	8,73	10,80	24%
1008	8,64	8,84	2%
1009	11,10	10,97	-1%
1010	8,93	10,74	20%
1011	10,63	12,96	22%
1012	8,31	12,28	48%
1013	9,99	11,98	20%
1014	8,89	10,70	20%
1015	8,20	9,23	12%
1016	8,72	9,94	14%
1017	10,15	11,71	15%
1018	8,39	8,36	0%
1019	9,18	8,72	-5%
1020	9,69	10,49	8%
1021	9,74	12,78	31%
1022	9,14	9,79	7%
1023	9,52	9,94	4%

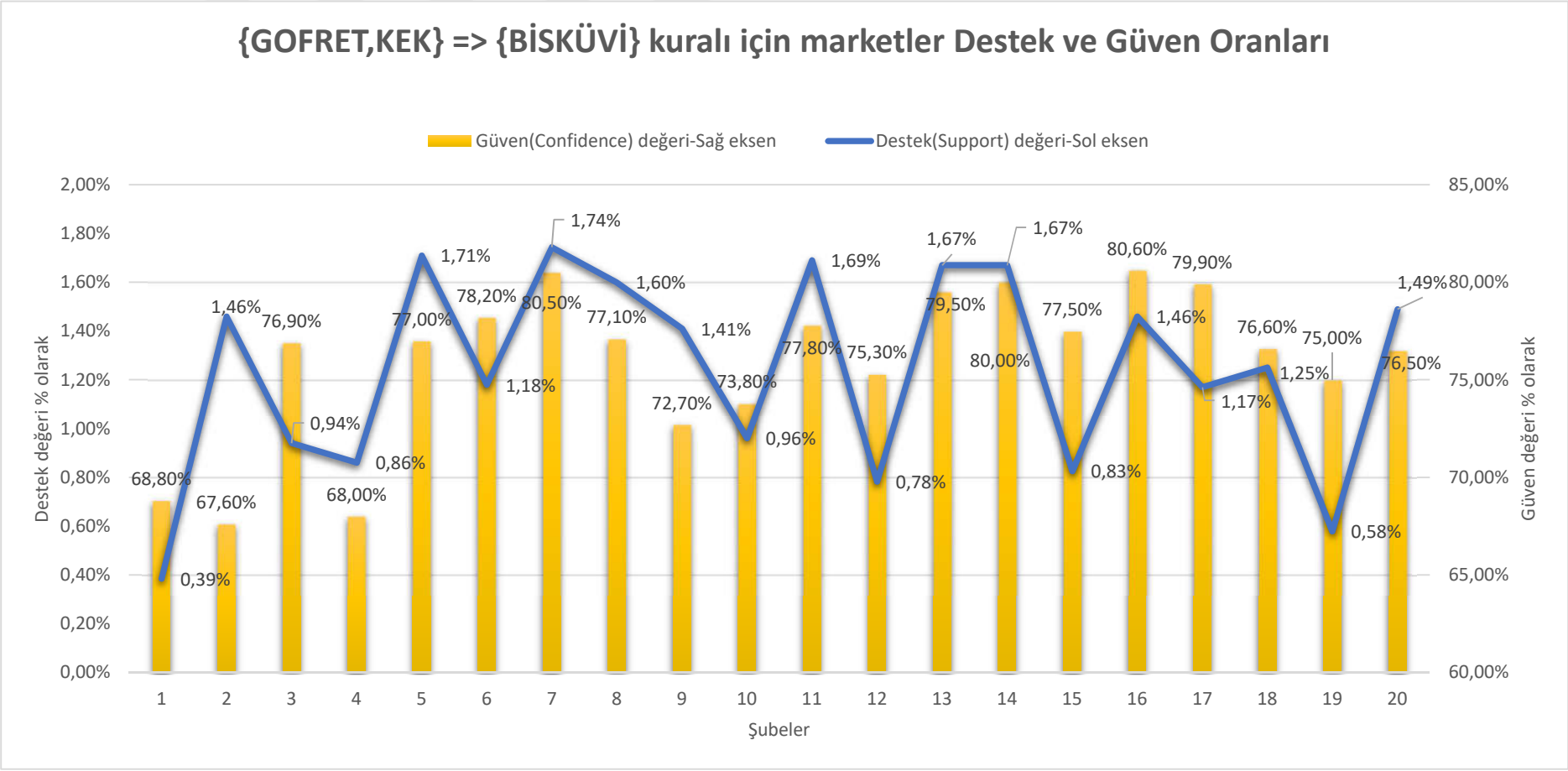
4.6. Analiz Sonuçlarından Elde Edilen Bazı Çıkarımlar

Birçok birliktelik kuralı ayrı ayrı veya çapraz olacak şekilde değerlendirilerek farklı müşteri profilleri elde edilebilir. Örnek olarak bu çalışmada gofret, Kek ve Bisküvi ürün gruplarının birlikte satın alan müşteriler üzerine bir analiz gerçekleştirilmiştir. Şekil 4.10 ile Şekil 4.24 arasındaki grafiklerde belirtildiği üzere bu alışveriş işlemini gerçekleştiren müşterilere ait bazı önemli bilgiler tespit edilmiştir. Bu bilgiler ışığında bazı çıkarımlar yapabilmek mümkündür. Bize yönlendirici bilgiler sunacak bazı örnek çıkarımlar aşağıda verilmiştir.

- İlk olarak belirlenen ürün gruplarına ait ürünleri birlikte tercih eden müşterilerin büyük çoğunluğunu 40-50 aralığındaki orta yaşlı müşteriler oluşturmaktadır. Bu oran şube 1003 de %66,81, şube 1004 de %39,27 olarak tespit edilmiştir.
- Müşterilerin çoğunluğu orta düzeyde(İlköğretim ve Lise) eğitim seviyesine sahiptir. Şube 1003 de %68,56, Şube 1004 de ise %73,59 olarak tespit edilmiştir.

- Düzenli bir geliri olmayan müşteri sayısı ise Şube 1003' de % 40,56 iken Şube 1004 için %31.1 olarak tespit edilmiştir. Düzenli geliri olmayan meslek gruplarına öğrenci, ev hanımı ve serbest meslek gruplarının verileri esas alınmıştır.
- Belirtilen alışverişi gerçekleştiren müşterilerin ezici çoğunluğunu evli olan erkek müşteriler oluşturmaktadır.
- Bu alışveriş işlemini şube 1003 de gerçekleştiren müşteriler en çok sonbahar dönemini tercih etmişlerdir. Bütün mevsimler içerisindeki oranı ise %31.74 dür. Şube 1004 de alışveriş yapan müşteriler ise İlkbaharı tercih etmişlerdir. Bu mevsimin oranı ise %30.04 olarak tespit edilmiştir.
- Haftanın hangi günlerinde daha çok tercih edildiğini inceleyecek olursak her iki şubede de belirtilen ürün gruplarının satışlarının diğer günlere nazaran hafta sonu daha yüksek miktarda gerçekleştiği tespit edilmiştir. Bu oranlar şube 1003'de %32.32, şube 1004 de ise %35.65 olarak bulunmuştur.
- Her iki şubede' de 17.00-21.00 saatleri aralığını kapsayan çeyrekte alışverişlerin büyük çoğunluğunun yapıldığı belirlenmiştir. Bu çeyrekte yapılan alışverişlerin şube 1003 için bütün alışverişler içerisindeki payı % 45.06 iken, Şube 1004 için bu oran %49.53'dür.
- Sepet tutarlarını incelediğimizde sepet tutarı 60 TL ve altında olan alışveriş sayısının oldukça fazla olduğu görülmektedir. Bu oran şube 1003 için %53.92, Şube 1004 için ise %33.47'dir. Bu verilerden yola çıkarak abur cubur diye adlandırabileceğimiz(Kek, Gofret ve Bisküvi ürün grupları) ürünleri şube 1003 den satın alan müşteriler küçük sepetlerde, şube 1004 ' den satın alan müşteriler ise büyük sepetlerde tercih etmiştir diyebiliriz.
- Belirtilen ürün gruplarının temel gıda maddelerini oluşturmadığı göz önüne alındığında alışveriş içerisindeki abur cubur ürünlerinin oranının yüksek olduğu market müşteri için daha yüksek gelir seviyesine sahip gibi bir çıkarım yapılabilir. Burada da Şekil 4.19'a bakacak olursak şube 1003 'ün şube 1004 'e göre biraz daha yüksek oranlara sahip olduğunu görebiliriz.
- Son olarak her iki marketin müşterilerine ait adres bilgileri incelendiğinde müşterilerin büyük çoğunluğunun alışveriş yaptıkları marketlere yakın bir konumda ikamet ettikleri tespit edilmiştir.

YSA kullanılarak gerçekleştirilen tahmin işlemlerine ait ayrıntılı bilgiler 6. bölümde belirtilmiştir.



Şekil 4.25. Farklı şubelerdeki destek ve güven değerleri

4.7. Farklı Şubelere Ait Birliktelik Kuralları

Şimdiye kadar yapılan çalışmalarda iki markete ait birliktelik kurallarından rastgele bir tanesi seçilerek bu kuralı sağlayan alışverişler ve bu alışverişi gerçekleştiren müşteriler analiz edilmiştir. Birçok şubesi olan bir firmanın doğru kampanyalar üretebilmesi için öncelikle doğru birliktelik kuralını seçmesi gerekmektedir. Daha sonra bu birliktelik kuralının hangi müşteri kitlesini etkilediğinin çıkarılması gerekmektedir. Bu sayede yapılacak kampanyaların asıl kitleleri tespit edilmiş olacaktır. Farklı şubelerde farklı birliktelik kuralları baskın çıkmış olabilir. Bu bizlere müşterilerin alışveriş tercihlerinin zenginliği olarak görülebilir ve doğru analiz edilmesi ile birlikte müşteri sadakatini artırıcı etkilerde bulunabilir. Çizelge 4.25. de rastgele seçilen 15 markete ait Güven değeri %80'in üstünde olan ve her marketin kendi içerisinde en yüksek destek değerine sahip birliktelik kuralı verilmiştir. Bu kurallar bir yıllık bir süre içerisinde belirtilen marketlerde alışveriş yapan bütün müşterilerin alışveriş tercihlerini temsil etmektedir. İsteğe bağlı olarak belirli bir kitle içinde birliktelik kuralları oluşturulacaktır. Buradan da görüldüğü üzere farklı marketlerdeki müşteri tercihleri benzerlik gösterdiği gibi ciddi farklılıklarda gösterebilmektedir.

Çizelge 4.25. Şubelere ait birliktelik kuralları

Şube	Kurallar	Destek Oranı	Güven Oranı
1005	{NARENCİYE,OTLAR-YEŞİLLİKLER,PATATES-SOĞAN-SARIMSAK} => {MEVSİM SEBZELERİ}	1,86%	81,70%
1006	{ÇAY,MAKARNA,SALÇA} => {ŞEKER-TATLANDIRICI}	0,91%	80,20%
1007	{AÇIK DANA ETİ,OTLAR-YEŞİLLİKLER,PATATES-SOĞAN-SARIMSAK} => {MEVSİM SEBZELERİ}	2,14%	85,80%
1008	{KEK,MAKARNA} => {BİSKÜVİ}	1,14%	80,10%
1009	{NARENCİYE,OTLAR-YEŞİLLİKLER,PATATES-SOĞAN-SARIMSAK} => {MEVSİM SEBZELERİ}	1,92%	81,40%
1010	{AÇIK DANA ETİ,PATATES-SOĞAN-SARIMSAK} => {MEVSİM SEBZELERİ}	2,15%	86,60%
1011	{GOFRET,KEK} => {BİSKÜVİ}	2,03%	82,00%
1013	{NARENCİYE,OTLAR-YEŞİLLİKLER,PATATES-SOĞAN-SARIMSAK} => {MEVSİM SEBZELERİ}	2,38%	82,50%
1014	{AÇIK DANA ETİ,OTLAR-YEŞİLLİKLER,PATATES-SOĞAN-SARIMSAK} => {MEVSİM SEBZELERİ}	2,25%	83,70%
1015	{AÇIK DANA ETİ,OTLAR-YEŞİLLİKLER,PATATES-SOĞAN-SARIMSAK} => {MEVSİM SEBZELERİ}	0,61%	83,10%
1016	{GOFRET,KEK} => {BİSKÜVİ}	0,92%	80,60%
1017	{AÇIK DANA ETİ,OTLAR-YEŞİLLİKLER,PATATES-SOĞAN-SARIMSAK} => {MEVSİM SEBZELERİ}	1,37%	85,60%
1018	{AÇIK DANA ETİ,OTLAR-YEŞİLLİKLER,PATATES-SOĞAN-SARIMSAK} => {MEVSİM SEBZELERİ}	1,36%	83,00%
1019	{OTLAR-YEŞİLLİKLER,PATATES-SOĞAN-SARIMSAK} => {MEVSİM SEBZELERİ}	10,32%	81,90%
1020	{AÇIK DANA ETİ,OTLAR-YEŞİLLİKLER,PATATES-SOĞAN-SARIMSAK} => {MEVSİM SEBZELERİ}	3,28%	84,80%

5. MÜŞTERİ BAZLI KAMPANYA OLUŞTURMA

Daha önceki bölümlerde müşterilere ait herhangi bir kısıtlama olmadan bütün marketler için ayrı ayrı birliktelik kuralları tespit edilmiştir. Daha sonrasında elde edilen bu birliktelik kurallarını sağlayan müşteri kitlesine ait bilgiler ortaya çıkarılmıştır. Bu bölümde ise belirli müşteri kitlelerine özgü yapılacak kampanyalar için önemli fikirler sağlayacak birliktelik kuralları oluşturulacaktır. Farklı parametrelere göre hedeflenen bir müşteri kitlesi için birliktelik kuralları üretmek mümkündür. Örnek ile açıklamak gerekirse öğretmenler gününde öğretmenlere yapılacak özel kampanya için öğretmenlerin alışveriş verilerinin analiz edilerek birliktelik kuralları oluşturulabilir.

5.1. Belirlenen Müşteri Kitlesine Özgü Oluşturulan Birliktelik Kuralları

Müşterilerin özelliklerine göre filtrelenerek oluşturulan bu kişiye özel oluşturulacak kampanya için müşterilerin hangi meslek grubuna dâhil olduğu, hangi eğitim seviyesinde olduğu, hangi zaman aralıklarında alışveriş yaptığı, alışverişlerindeki ortalama satın alma gücü, satın alma zamanı vb.. bir çok parametreye göre birliktelik kuralları üretilebilir.

Birkaç örnek çalışma hazırlanmıştır. Bu örnek çalışmanın hedef kitleleri sırası ile ;

- Meslek grubu Doktor olan müşteriler
- Meslek grubu Öğretmen olan müşteriler
- Meslek grubu Ev hanımı olan müşteriler
- Meslek grubu Öğrenci olan müşteriler
- Öğrenim Durumu Yüksek lisans olan müşteriler
- Belirli bir tarih hedeflenerek, son iki hafta içerisinde alışveriş yapmış olan müşterilere özel birliktelik kuralları analiz edilmiştir.

❖ Meslek Grubu Doktor olan müşterilerin 1. ve 2. yıllarındaki alışverişleri incelenerek elde edilen birliktelik kuralları Çizelge 5.1. ve Çizelge 5.2. de verilmiştir. (Bu bölümdeki bütün çizelgelerde ilgili yıllardaki en yüksek destek oranına sahip ilk 2 kural verilmiştir.)

Çizelge 5.1. 1. Yıl doktor olan müşteriler için birliktelik kuralları

Birliktelik Kuralları	Destek Oranı	Güven Oranı
{NORMAL YUMURTALAR,OTLAR-YEŞİLLİKLER} => {MEVSİM SEBZELERİ}	4,69%	81,60%
{EKMEK,OTLAR-YEŞİLLİKLER} => {MEVSİM SEBZELERİ}	4,24%	80,00%

Çizelge 5.2. 2. Yıl doktor müşteriler için birliktelik kuralları

Birliktelik Kuralları	Destek Oranı	Güven Oranı
{AÇIK DANA ETİ,OTLAR-YEŞİLLİKLER,PATATES-SOĞAN-SARIMSAK} => {MEVSİM SEBZELERİ}	1,62%	80,50%
{NARENCİYE,OTLAR-YEŞİLLİKLER,PATATES-SOĞAN-SARIMSAK} => {MEVSİM SEBZELERİ}	1,45%	81,20%

- Meslek Grubu Öğretmen olan müşterilerin 1. ve 2. yıllarındaki alışverişleri incelenerek elde edilen birliktelik kuralları Çizelge 5.3. ve Çizelge 5.4. de verilmiştir.

Çizelge 5.3. 1. Yıl öğretmen olan müşteriler için birliktelik kuralları

Birliktelik Kuralları	Destek Oranı	Güven Oranı
{AÇIK DANA ETİ,OTLAR-YEŞİLLİKLER,PATATES-SOĞAN-SARIMSAK} => {MEVSİM SEBZELERİ}	1,62%	80,50%
{NARENCİYE,OTLAR-YEŞİLLİKLER,PATATES-SOĞAN-SARIMSAK} => {MEVSİM SEBZELERİ}	1,45%	81,20%

Çizelge 5.4. 2. Yıl öğretmen olan müşteriler için birliktelik kuralları

Birliktelik Kuralları	Destek Oranı	Güven Oranı
{AÇIK DANA ETİ,OTLAR-YEŞİLLİKLER,PATATES-SOĞAN-SARIMSAK} => {MEVSİM SEBZELERİ}	1,81%	81,30%
{GOFRET,KEK} => {BİSKÜVİ}	1,40%	81,50%

- Meslek Grubu Ev hanımı olan müşterilerin 1. ve 2. yıllardaki alışverişleri incelenerek elde edilen birliktelik kuralları Çizelge 5.5. ve Çizelge 5.6. da verilmiştir.

Çizelge 5.5. 1. Yıl ev hanımı olan müşteriler için birliktelik kuralları

Birliktelik Kuralları	Destek Oranı	Güven Oranı
{ÇAY,DİĞER BAKLİYAT,MAKARNA,PİRİNÇ} => {ŞEKER-TATLANDIRICI}	0,56%	80,00%
{AYÇİÇEK YAĞLARI,ÇAY,TUZ} => {ŞEKER-TATLANDIRICI}	0,54%	80,50%

Çizelge 5.6. 2. Yıl ev hanımı olan müşteriler için birliktelik kuralları

Birliktelik Kuralları	Destek Oranı	Güven Oranı
{KRAKER} => {BİSKÜVİ}	0,68%	80,10%
{ÇAY,PASTA MALZEMELERİ,PİRİNÇ} => {MAKARNA}	0,63%	80,70%

- Meslek Grubu Öğrenci olan müşterilerin 1. ve 2. yıllardaki alışverişleri incelenerek elde edilen birliktelik kuralları Çizelge 5.7. ve Çizelge 5.8.'de verilmiştir.

Çizelge 5.7. 1. Yıl öğrenci olan müşteriler için birliktelik kuralları

Birliktelik Kuralları	Destek Oranı	Güven Oranı
{GOFRET,KEK} => {BİSKÜVİ}	1,79%	81,90%
{AÇIK DANA ETİ,OTLAR-YEŞİLLİKLER,PATATES-SOĞAN-SARIMSAK} => {MEVSİM SEBZELERİ}	1,31%	86,70%

Çizelge 5.8. 2. Yıl öğrenci olan müşterileri için birliktelik kuralları

Birliktelik Kuralları	Destek Oranı	Güven Oranı
{AÇIK DANA ETİ,OTLAR-YEŞİLLİKLER,PATATES-SOĞAN-SARIMSAK} => {MEVSİM SEBZELERİ}	1,57%	87,30%
{PİRİNÇ,TUZ} => {MAKARNA}	1,21%	83,30%

- Öğrenim durumu yüksek lisans olan müşterilerin 1. ve 2. yıllarındaki alışverişleri incelenerek elde edilen birliktelik kuralları Çizelge 5.9. ve Çizelge 5.10. de verilmiştir.

Çizelge 5.9. 1. Yılı yüksek lisans mezunu olan müşteriler için birliktelik kuralları

Birliktelik Kuralları	Destek Oranı	Güven Oranı
{NARENCİYE,OTLAR-YEŞİLLİKLER,PATATES-SOĞAN-SARIMSAK} => {MEVSİM SEBZELERİ}	2,41%	83,80%
{AÇIK DANA ETİ,OTLAR-YEŞİLLİKLER,PATATES-SOĞAN-SARIMSAK} => {MEVSİM SEBZELERİ}	2,34%	85,40%

Çizelge 5.10. 2. Yıl yüksek lisans mezunu olan müşteriler için birliktelik kuralları

Birliktelik Kuralları	Destek Oranı	Güven Oranı
{OTLAR-YEŞİLLİKLER,PATATES-SOĞAN-SARIMSAK} => {MEVSİM SEBZELERİ}	6,92%	86,60%
{NARENCİYE,OTLAR-YEŞİLLİKLER} => {MEVSİM SEBZELERİ}	4,65%	87,00%

- Son iki hafta içinde alışveriş yapmış olan müşterilerin 1. ve 2. yıllardaki alışverişleri incelenerek elde edilen birliktelik kuralları Çizelge 5.11 ve Çizelge 5.12 de verilmiştir. Son iki haftalık süre hesaplanırken her iki yıl içinde yılbaşı tarihi(1 Ocak) temel alınmıştır.

Çizelge 5.11. 1. Yıl itibari ile belirlenen tarih aralığında alışveriş yapan müşterilere ait birliktelik kuralları

Birliktelik Kuralları	Destek Oranı	Güven Oranı
{GOFRET,KEK} => {BİSKÜVİ}	1,62%	82,70%
{AÇIK DANA ETİ,OTLAR-YEŞİLLİKLER,PATATES-SOĞAN-SARIMSAK} => {MEVSİM SEBZELERİ}	1,18%	81,60%

Çizelge 5.12. 2. Yıl itibari ile belirtilen tarih aralığında alışveriş yapan müşterilere ait birliktelik kuralları

Birliktelik Kuralları	Destek Oranı	Güven Oranı
{NARENCİYE,OTLAR-YEŞİLLİKLER,PATATES-SOĞAN-SARIMSAK} => {MEVSİM SEBZELERİ}	1,69%	81,80%
{AÇIK DANA ETİ,OTLAR-YEŞİLLİKLER,PATATES-SOĞAN-SARIMSAK} => {MEVSİM SEBZELERİ}	1,16%	80,90%

5.2. Belirlenen Birliktelik Kurallarına Özgü Müşteri Profili Tespiti

Marketçilik sektöründe faaliyet gösteren firmalar sattığı ürünleri paket olarak müşterileri için sunarak, onlara satın alacakları ürünleri hem daha uygun fiyata hem de daha kolay ulaşılabilir olmasını sağlamışlardır. Bu bölümdeki çalışmamızda Ramazan ayında müşterilere sunulmak üzere örnek bir sepet uygulaması yapılmıştır ve bu hazırlanan sepetin hangi müşteri kitlesi tarafından rağbet göreceği üzerine bir müşteri analizi çalışması bu veriler ışığında yapılmıştır.

İlk olarak 1. ve 2. senelerdeki ramazan ayında müşterilerin alışveriş bilgileri analiz edilmiştir. Bu analiz neticesinde en çok birlikte satın alınan Kod_3 diye adlandırdığımız ürün grupları tespit edilmiştir. Bu elde ettiğimiz birliktelik kuralları her iki sene içerisindeki ramazan ayında yapılan satışları kapsamaktadır. 1.yıl ve 2.yıl ramazan ayı için oluşturulan birliktelik kuralları Çizelge 5.13.'de ve Çizelge 5.14.'de sırası ile verilmiştir.

Çizelge 5.13. 1. yıl Ramazan ayı birliktelik kuralları

BİRLİKTELİK KURALLARI	DESTEK ORANI	GÜVEN ORANI
{OTLAR-YEŞİLLİKLER,PATATES-SOĞAN-SARIMSAK} => {MEVSİM SEBZELERİ}	3,24%	74,60%
{AÇIK DANA ETİ,PATATES-SOĞAN-SARIMSAK} => {MEVSİM SEBZELERİ}	2,83%	71,60%
{AÇIK DANA ETİ,OTLAR-YEŞİLLİKLER} => {MEVSİM SEBZELERİ}	2,32%	72,50%
{AÇIK DANA ETİ,OTLAR-YEŞİLLİKLER,PATATES-SOĞAN-SARIMSAK} => {MEVSİM SEBZELERİ}	1,58%	83,10%
{DİĞER MEYVELER,PATATES-SOĞAN-SARIMSAK} => {MEVSİM SEBZELERİ}	1,40%	74,40%
{PATATES-SOĞAN-SARIMSAK,TEREYAĞI} => {MEVSİM SEBZELERİ}	1,33%	73,60%
{DİĞER MEYVELER,OTLAR-YEŞİLLİKLER} => {MEVSİM SEBZELERİ}	1,31%	77,10%
{KAVUN-KARPUZ,PATATES-SOĞAN-SARIMSAK} => {MEVSİM SEBZELERİ}	1,16%	73,10%
{KETÇAP-HAZIR SOSLAR,PATATES-SOĞAN-SARIMSAK} => {MEVSİM SEBZELERİ}	1,14%	87,00%
{KETÇAP-HAZIR SOSLAR,MEVSİM SEBZELERİ} => {PATATES-SOĞAN-SARIMSAK}	1,14%	74,20%

Çizelge 5.14. 2. Yıl Ramazan ayı birliktelik kuralları

BİRLİKTELİK KURALLARI	DESTEK ORANI	GÜVEN ORANI
{OTLAR-YEŞİLLİKLER,PATATES-SOĞAN-SARIMSAK} => {MEVSİM SEBZELERİ}	2,98%	81,60%
{AÇIK DANA ETİ,PATATES-SOĞAN-SARIMSAK} => {MEVSİM SEBZELERİ}	2,93%	77,20%
{AÇIK DANA ETİ,OTLAR-YEŞİLLİKLER} => {MEVSİM SEBZELERİ}	2,54%	72,10%
{AÇIK DANA ETİ,OTLAR-YEŞİLLİKLER,PATATES-SOĞAN-SARIMSAK} => {MEVSİM SEBZELERİ}	1,60%	87,00%
{PATATES-SOĞAN-SARIMSAK,TEREYAĞI} => {MEVSİM SEBZELERİ}	1,30%	80,90%
{KETÇAP-HAZIR SOSLAR,PATATES-SOĞAN-SARIMSAK} => {MEVSİM SEBZELERİ}	1,16%	91,40%
{KETÇAP-HAZIR SOSLAR,MEVSİM SEBZELERİ} => {PATATES-SOĞAN-SARIMSAK}	1,16%	75,90%
{GAZOZ,PATATES-SOĞAN-SARIMSAK} => {MEVSİM SEBZELERİ}	1,13%	73,70%
{AÇIK KUZU ETİ,PATATES-SOĞAN-SARIMSAK} => {MEVSİM SEBZELERİ}	1,07%	83,30%
{NARENCİYE,OTLAR-YEŞİLLİKLER} => {MEVSİM SEBZELERİ}	1,05%	78,80%

1. ve 2. yıllar için tespit edilen kuralları incelendiğinde her 2 yıl içinde benzer bir kuralın en yüksek destek değerine sahip olduğu görülmektedir. Yani müşteriler hem 1.

senenin hem de 2. senenin ramazan ayın da benzer ürün gruplarına dâhil olan ürünleri tercih etmişlerdir. Bu kural ilgili çizelgelerde kırmızı renk ile belirtilmiştir. {OTLAR-YEŞİLLİKLER, PATATES-SOĞAN-SARIMSAK} => {MEVSİM SEBZELERİ} kuralı 1. yıl için destek değeri %3,24 güven değeri %74,60 ve 2 Yıl için destek değeri %2,98 güven değeri %81,60 olarak tespit edilmiştir. İlk aşamada tespit edilen bu kurallar sepetin hazırlanmasında fikir verici olup yetersiz kalmaktadır. İkinci aşamada bu kuralın sağlandığı alışverişler için ürün odaklı birliktelik kuralları çıkarılmıştır. Müşterilerin (OTLAR-YEŞİLLİKLER, PATATES-SOĞAN-SARIMSAK ve MEVSİM SEBZELERİ) ürün grupları içerisinde hangi ürünleri birlikte tercih edildiğine dair bilgiler ortaya çıkarılmıştır. 1. yıl için bu ürünler Çizelge 5.15 de verilmişken, 2 yıl için Çizelge 5.16 da verilmiştir.

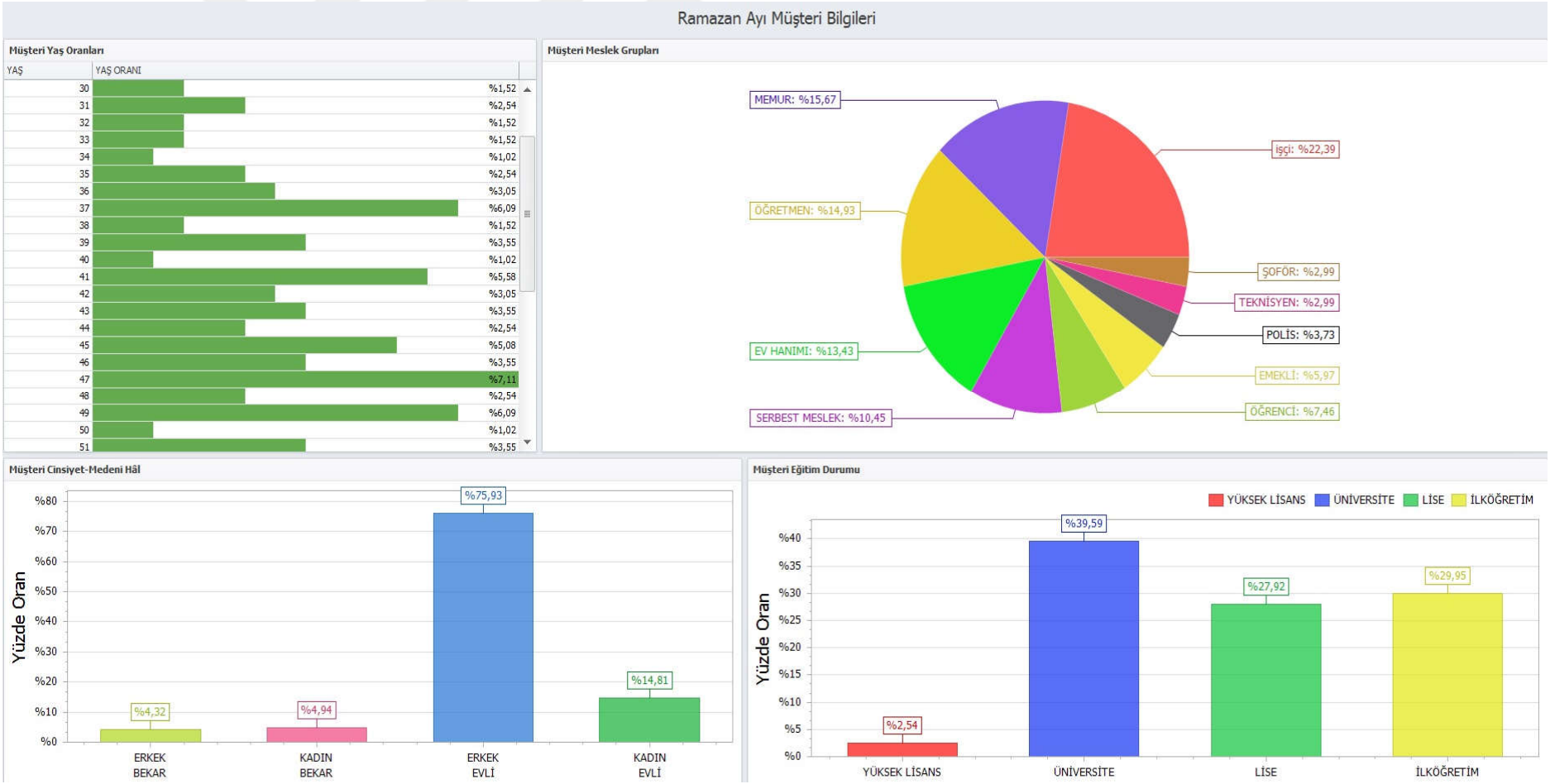
Çizelge 5.15. 1 yıl belirtilen ürün gruplarındaki ürünlere ait birliktelik kurallar

Birliktelik Kuralları	Destek Oranı	Güven Oranı
{Kırmızı Biber, Kuru Soğan} => {Maydanoz}	2,54%	60,20%
{ Maydanoz, Kırmızı Biber } => { Kuru Soğan }	2,54%	65,70%
{Çarliston Biber, Kırmızı Biber } => { Kuru Soğan }	1,93%	62,50%
{ Maydanoz, Sarımsak} => { Kuru Soğan }	1,47%	66,50%
{ Kırmızı Biber, Sarımsak } => { Kuru Soğan }	1,41%	72,70%

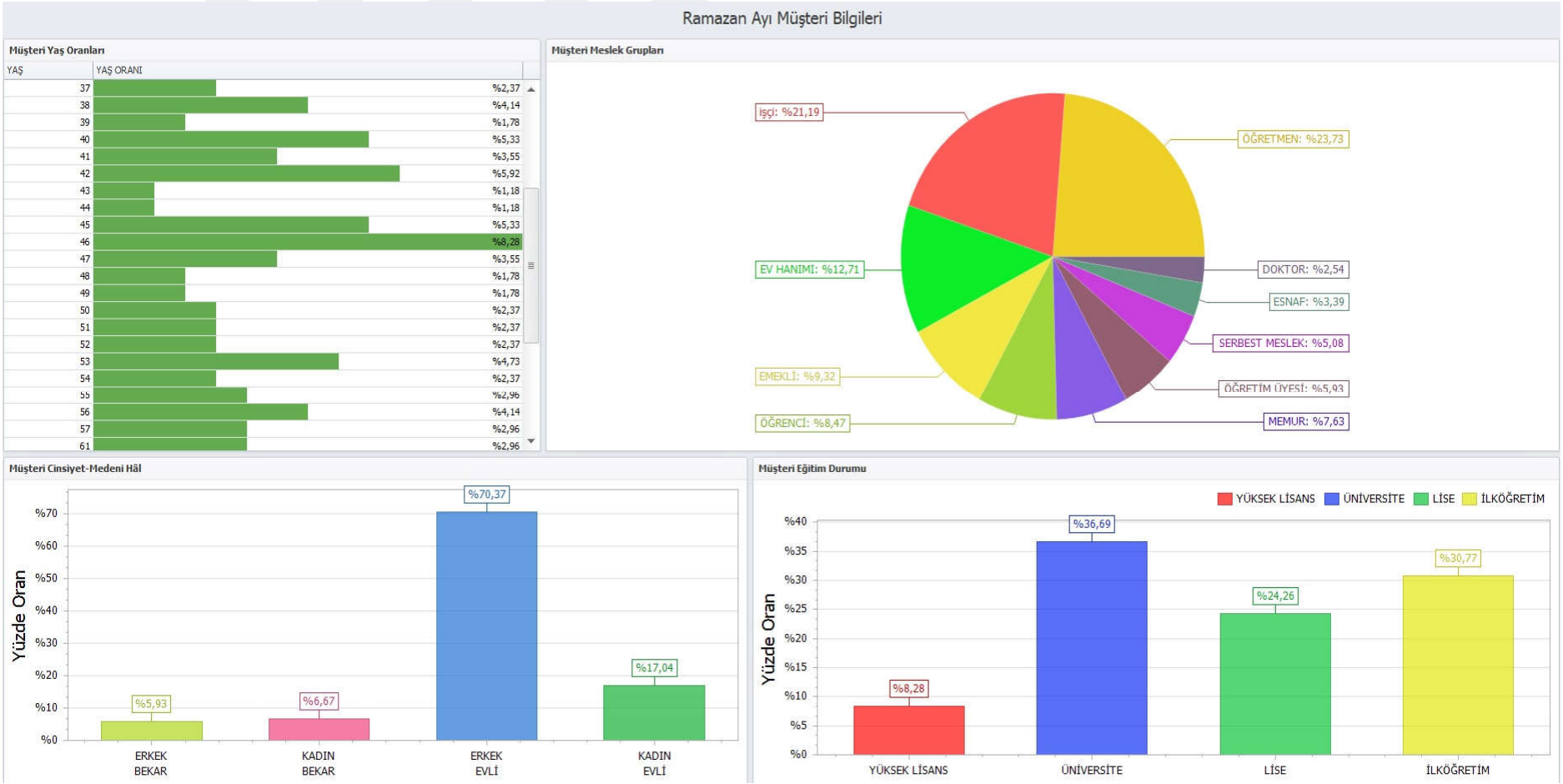
Çizelge 5.16. 2. yıl belirtilen ürün gruplarındaki ürünlere ait birliktelik kuralları

Birliktelik Kuralları	Destek Oranı	Güven Oranı
{ Maydanoz, Kırmızı Biber } => { Kuru Soğan }	3,15%	74,40%
{ Kırmızı Biber, Kuru Soğan } => { Maydanoz }	3,15%	61,90%
{ Çarliston Biber, Kırmızı Biber} => { Kuru Soğan }	2,38%	64,50%
{Yerli Domates, Kırmızı Biber } => { Kuru Soğan }	2,33%	63,10%
{ Maydanoz, Sarımsak } => { Kuru Soğan }	1,96%	71,90%

Elde edilen birliktelik kuralları incelendiğinde her iki yıl için de Kırmızı Biber, Kuru Soğan ve Maydanoz ürünlerinin bulunduğu birliktelik kuralının en yüksek destek değerine sahip olduğu görülmektedir. Bu birliktelik kuralına dayanarak hazırlanacak ramazan sepeti için bu üçlü ürün grubunun birlikte paket olarak sunulması sağlanabilir. Bu paket ürünler hem kırmızı biber, maydanoz ve kuru soğan ürünlerini tercih eden müşterilere hem de maydanoz ve kırmızı biber ürünlerini tercih ettikten sonra kuru soğan satın almayan müşterilere tavsiye olarak sunulabilir. Temel hedeflerden biriside ilk iki ürünü benzer olarak tercih eden müşteriler için kampanya sayesinde kuru soğan satın almayan müşterileri bu ürünü satın almaya teşvik etmek. Bu şekilde birbiri ile bağlantılı başka ürünlerinde sepete eklenmesi ile sepete zenginlik katılabilir. 1. yıl ramazan ayında (Kırmızı Biber, Kuru Soğan ve Maydanoz) ürünlerini sepetinde bulunduran müşterilere ait bazı özellikler Şekil 5.1 de verilmiştir. Yine aynı koşulları sağlayan müşterilerin 2. yıla ait bilgileri ise Şekil 5.2 de verilmiştir. Bu bilgilerin değerlendirilmesi ile hazırlanan sepetin ilgili müşterilere sunulması sağlanacaktır.



Şekil 5.1 1. Yıl belirtilen ürünler için ramazan ayındaki müşteri bilgileri



Şekil 5.2 2. Yıl belirtilen ürünler için ramazan ayındaki müşteri bilgileri

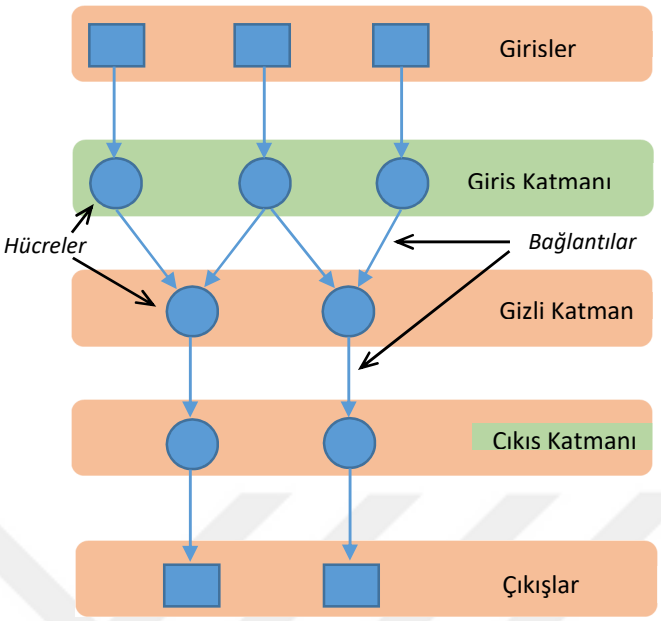
6. BİRLİKTELİK KURALLARI İLE TESPİT EDİLEN KAMPANYA ÜRÜNLERİNİN YSA 'İLE SATIŞ MİKTARLARININ TAHMİN EDİLMESİ

İnsan beyninin modellenmesi ile tasarlanan yapay sinir ağları, beyinde bulunan biyolojik sinir hücrelerine benzer özelliklerde bulunan yüksek miktardaki sinir hücresinden oluşmaktadır. Biyolojik sinir hücrelerine benzer bir şekilde yapay sinir hücrelerinin yapısında aldıkları giriş sinyalini, toplayıp işledikleri ve çıktı olarak ilettikleri bölümlerden oluşmaktadır. 5 katmandan oluşan YSH'nın yapısı aşağıdaki şekilde sıralanabilir; 1.Girdiler, 2. Ağırlıklar, 3.Toplama Fonksiyonu, 4. Aktivasyon Fonksiyonu, 5. Çıktılar [59,60]. Girdiler yapay sinir hücrelerine dış dünyadan gelebileceği gibi başka bir hücreden de gelebilir. YSH'a gelen bilgiler geldikleri bağlantılara ait ağırlıklar ile çarpılarak çekirdeğe iletilir. Bu sayede girdilerin çıktılar üzerindeki etkisine müdahale etme imkânı oluşturulmuştur. Ağırlıklar değişken veya sabit değerlerden oluşabilir [61]. Yapay sinir hücresine ait ağırlıklar ile çarpılarak gelen girdileri toplayıp, hücrenin net girdisini hesaplayan fonksiyona toplama fonkisyonu denilmektedir. Farklı problem türleri değişik toplama fonksiyonlarında değişik oranlarda başarılı sonuçlar verebilir. Bir çok toplama fonksiyonu içerisinde en çok kullanılan fonksiyon gelen girdi değeri ile sahip olduğu ağırlık değerinin çarpılarak toplanması ile gerçekleştirilen net girdi değerinin hesaplandığı yöntemdir.

$$\boldsymbol{NET = \sum_{i=1}^{n} G_i A_i} \qquad (6,1)$$

Denklem (6,1) de belirtilen formüldeki G_i girişi, A_i ağırlıkları ve n ise hücreye gelen toplam girdi sayısını göstermektedir [62]. Net girdinin hesaplanarak hücredeki girdiye karşılık üreteceği çıktıyı oluşturan fonksiyona aktivasyon fonksiyonu denilmektedir. Çeşitli aktivasyon fonksiyonları bulunmaktadır. Bunlar içerisinde yaygın olarak kullanılanlarından bazıları basamak fonksiyonu, sigmoid fonksiyon, hiperbolik tanjant fonksiyonu olarak verilebilir [61]. YSH'nın çıktı değerini aktivasyon fonksiyonu belirlemektedir. Bu çıktı değeri dış dünyaya verilebileceği gibi tekrar ağ içerisine dahil edilerek kullanılabilir [59]. Şekil 6.1'de yapay sinir hücreleri tarafından oluşturulan yapay sinir ağının genel yapısı görülmektedir. Şekilde belirtilen yapay sinir ağı giriş, çıkış ve 3 farklı katmanı sisteminde bulundurmaktadır. Giriş katmanı giriş verilerinin ağa sunulduğu katmandır. Bu katman üzerinden veriler herhangi bir işleme tabi tutulmadan gizli katmana iletilir. Gizli katman ağ üzerinde temel işlevlerin gerçekleştirildiği ana katmandır. Ağ içerisinde birden fazla gizli katman bulunabilir. Problemin büyüklüğüne ve türüne göre gizli katman içerisindeki nöron sayıları değişmektedir. Bu tabaka girdi olarak gelen veriyi uygun fonksiyon ile işleyerek çıkış katmanına aktarmaktadır. Çıkış katmanı ağın en uç

katmanı olarak görev yapar ve işlemler tamamlandıktan sonraki çıktıyı dışarı aktarmaktadır [63].



Şekil 6.1. Genel YSA yapısı

Bu bölümde birbirleri arasında güçlü birliktelik bağları bulunan ürünlerinin satış miktarlarının önceden tespit edilmesine yönelik bir çalışma yapay sinir ağları tekniği uygulanarak yapılmıştır. Bu sayede birliktelik kuralları göz önüne alınarak oluşturulan bir kampanyanın henüz fikir aşamasında iken yapılan bu yüksek doğruluk payına sahip tahmin işlemleri ile hangi markette bu belirtilen ürün gruplarının müşteriler tarafından ne kadar rağbet göreceği hakkında bilgiler elde edilmiştir. Bu uygulamada belirlenen ürünlerin 3 farklı marketteki 34 aylık satış verileri kullanılmıştır. Bu verilerden ilk 32 aylık bölümü yapay sinir ağının eğitilmesi için kullanılmıştır. Kalan 2 aylık satış verisi ise tahmin edilen satış verilerinin karşılaştırılması ve başarı oranının tespit edilmesinde kullanılmaktadır. Birliktelik kuralları istenen tarih, şube ve ürün grupları kısıtlarına göre ayrıntılı olarak oluşturulabilmektedir. Bu bölümdeki çalışmamızda birliktelik kurallarının üretilmesinde kullanılacak veriler bir önceki bölümde belirtildiği üzere 2. yılda ramazan ayında satışı yapılmış manav ürünlerine ait verilerdir. Bu veriler ışığında oluşturulan birliktelik kurallarından en yüksek destek oranına sahip ilk 5 kural Çizelge 6.1. de verilmiştir. Bu belirlenen kurallardan Çizelge 6.1 e kırmızı renkte gösterilen maydanoz, kuru soğan, kırmızı biber ürünlerinin birlikte paket olarak satıldığı ürünlerin satış miktarı tahminleri yapılmıştır.

Çizelge 6.1 Birliktelik kuralları

Birliktelik Kuralları	Destek Oranı	Güven Oranı
{ Maydanoz, Kırmızı Biber } => { Kuru Soğan }	3,15%	74,40%
{ Kırmızı Biber, Kuru Soğan } => { Maydanoz }	3,15%	61,90%
{ Çarliston Biber, Kırmızı Biber} => { Kuru Soğan }	2,38%	64,50%
{Yerli Domates, Kırmızı Biber } => { Kuru Soğan }	2,33%	63,10%
{ Maydanoz, Sarımsak } => { Kuru Soğan }	1,96%	71,90%

Tespit edilen Birliktelik kuralına göre maydanoz ve kırmızı biber satın alan müşterilerin aynı sepette kuru soğan tercih etme oranı %74.4 dür. Aynı zamanda Maydanoz, kırmızı biber ve kuru soğan ürünlerinin beraber satıldığı alışveriş miktarlarının bütün alışverişler içerisindeki oranı %3.15 dir. Bu satış fiyatları gibi bilgiler ile YSA eğitilerek son iki ay içerisindeki satış miktarı tahmin edilecektir.

Yapay sinir ağı eğitiminde 4 giriş ve 1 çıkış parametresi kullanılmıştır. Giriş parametreleri "ürünlerin ortalama fiyatları", "alışveriş yapılan market", "satış yapılan ay", "satış yapılan yıl" çıkış parametresi olarak ise "ürünlerin satış miktarı" seçilmiştir. Tahmin işlemi için belirlenen ürünlere ait ortalama satış fiyatı, satış miktarı, satıldığı market ve satıldığı tarih bilgileri tespit edilmiştir. Çizelge 6.2 de örnek olarak 34 aylık verinin 3 aylık kısmı verilmiştir. Çizelge incelendiğinde 1. senenin 3. Ayında belirlenen 3 ürünün birlikte olacak şekilde market 1001'de 53 adet satılmıştır ve bu ürün grubunun paket olarak ortalama fiyatı 7,9342'TL dir.

Çizelge 6.2. Ürün kümesi satış verileri

Satış Miktarı	Ortalama Fiyat	Market	Ay	Yıl
53	7,9342	1001	3	1.yıl
51	7,9799	1002	3	1. yıl
48	7,9889	1003	3	1. yıl
29	8,9131	1001	4	1. yıl
27	8,9788	1002	4	1. yıl
47	8,9745	1003	4	1. yıl
24	7,0587	1001	5	1. yıl
20	7,2028	1002	5	1. yıl
33	7,4244	1003	5	1. yıl

Tasarlanan YSA'nın bu veriler ışığında eğitilmesi ve tahmin sonuçlarının belirlenmesi aşamaları MATLAB yazılımının NNTOOL aracı ile yapılmıştır. MATLAB yazılımında bu verilerin kullanılabilmesi için bütün verilerin normalize edilerek 0-1 aralığındaki değerlere dönüştürülmesi gerekmektedir. Örnek olarak Çizelge 6.2. deki verilerin normalize edilmiş hali Çizelge 6.3. de verilmiştir.

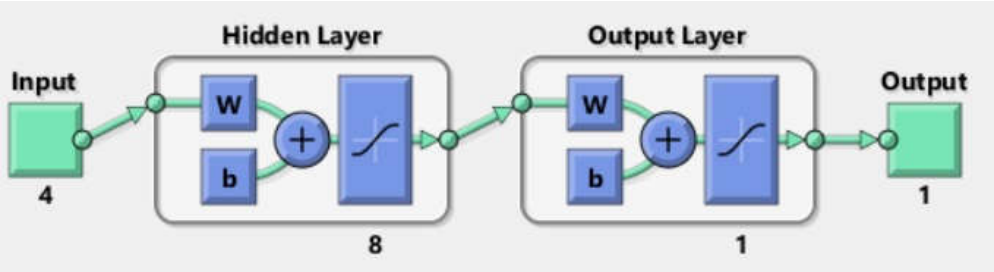
Çizelge 6.3. Ürün kümesi satış verileri normalize edilmiş hali

Satış miktarı	Ortalama Fiyat	Market	Ay	Yıl
0,519095789	0,724589532	0	0,181818	0
0,5	0,731924182	0,5	0,181818	0
0,464760528	0,733368642	1	0,181818	0
0,26587531	0,881698686	0	0,272727	0
0,244777044	0,892243247	0,5	0,272727	0
0,455367465	0,891553116	1	0,272727	0
0,222130471	0,584075626	0	0,363636	0
0,172089182	0,607203043	0,5	0,363636	0
0,308608588	0,64276887	1	0,363636	0

Aşağıda bu çalışmada bütün veri türleri uygulanan min-max normalizasyon işleminin formülü (6,2) numaralı denklemde belirtilmiştir.

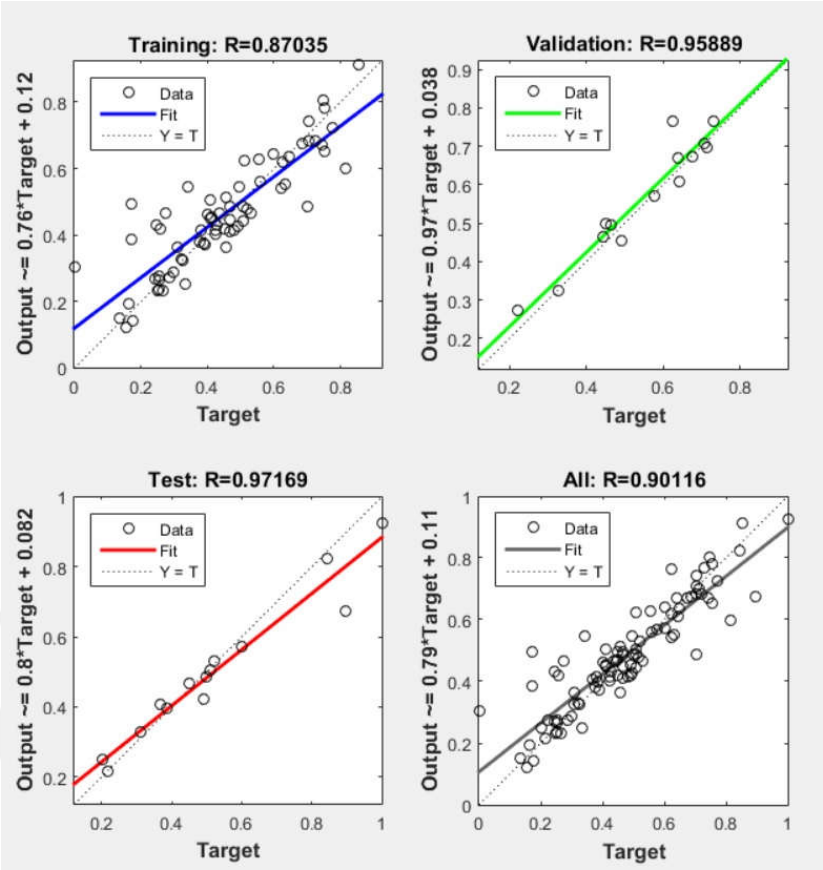
$$X' \frac{X - Xmin}{Xmax - Xmin} \qquad (6,2)$$

Hazırlanan veriler tasarlanan YSA üzerinde uygulanarak öncelikle ağın eğitilmesi işlemi gerçekleştirilmiştir. YSA'nın eğitim sürecinde optimum başarıyı sağlamak için ağ yapısı üzerinde çeşitli alternatifler üzerinde çalışılmış ve Şekil 6.2. de görülen ağ yapısı oluşturulmuştur. Gerçekleşen eğitim sürecinde oluşturulan regresyon eğrileri ile tahmin için en uygun olan eğitimli ağın tespit edilmesi sağlanmıştır. Yapılan denemeler sonucunda en uygun YSA yapısı 4 giriş, 8 orta katman ve 1 çıkış nöronundan oluştuğu tespit edilmiştir.



Şekil 6.2. Tasarlanan yapay sinir ağı

Yapay sinir ağı üzerinde gerçekleştirilen başarılı eğitim ve tahmin sürecini açıklayan regresyon eğrileri Şekil 6.3. de verilmiştir. Regresyon eğrileri sayesinde ağın ne kadar sağlıklı eğitildiğini ve test tahminlerinin başarı oranını görebilmekteyiz.



Şekil 6.3. Regresyon eğrileri

Uygun yapının sağlanması ile veriler analiz işlemine sokulmuştur. 3 Market için 32 aylık satış verisi kullanılarak eğitilen YSA'nın bu marketler için son 2 aylık tahmin verileri ve başarı oranı Çizelge 6.4. de görüldüğü üzere oldukça başarılı bir şekilde gerçekleşmiştir.

Çizelge 6.4. Tahmin sonuçları

Market No	Ay	Tahmin Edilen Satış Miktarı	Gerçek Satış Miktarı	Tahmin Başarı Oranı
1001	11	52,03	45	86,49%
1002	11	61,19	57	93,15%
1003	11	33,00	31	93,93%
1001	12	55,29	49	88,63%
1002	12	62,07	67	92,05%
1003	12	21,38	25	83,05%
		284,96	274	96,15%

Yapılan tahmin sonuçlarında birliktelik kurallarına göre oluşturulan paket ürünlerinin satış ve tahmin bilgileri verilmiştir. Bu tahmin sonuçlarına göre 11. Ay da 1001 numaralı şubede belirlenen paket ürünlerden 45 adet satılmışken, Ysa ile belirlenen tahmin değerinde ise yaklaşık 52 adet satış yapabileceği sonucuna ulaşılmıştır. Buradan yaklaşık

%86.5 başarı oranında tahmin işleminin gerçekleştirildiği sonucuna ulaşılmıştır. Bir başka örnek'de ise 12. ayda 1002 numaralı şubede bu paket ürünlerden 67 adet satılmıştır ve tahmin sonucu olarak ise yaklaşık 62 adet satış yapabileceği bilgisi tespit edilmiştir. Buradaki tahmin sonucu %92 başarı ile tespit edilmiştir.3 market ve 2 aylık satış verileri kümülatif olarak incelendiğinde ise gerçekleşen satış miktarı 274 adet gerçekleşmesine rağmen tahmini satış miktarı yaklaşık 285 olarak tespit edilmiştir. Bu durumda satış tahmini % 96.15 başarı oranı ile gerçekleştirilmiş oldu.

7. TARTIŞMA VE SONUÇ

Bu tez çalışmamızda market sektöründe üretilen bazı satış ve müşteri verilerinin daha anlamlı sonuçlara dönüştürülmesi için çeşitli analiz uygulamaları yapılmıştır. Bu uygulama gerçek bir sistem üzerinde uygulanabilir şekilde hazırlanmıştır. Firma ve müşteri avantajları gözetilerek gerçekleştirilmiştir. Temel amaçlardan birisi firmanın müşterilerin sadakatini artırıcı kampanyalar yapabilmesi ve satışları artırabilmesi için değeri sonuçlar elde etmektir. Genel kampanyalar yapmaktansa, müşterilerin alışveriş tercihlerinin belirlenmesi ile her müşterinin ihtiyacı olan ürünleri müşteriye özel kampanyalarla ulaştırılması hedeflenmiştir. Müşterilerin düzenli olarak satın aldıkları ürünleri hangi zamanlarda gerçekleştirdiği bilgisinin tespit edilmesi sayesinde bir sonraki alışveriş gerçekleşmeden ona özel yapılacak kampanya ile firmaya olan bağlılığını artırmak temel amaçlar arasında sayılabilir. Firmanın sadık müşterilerinin artması, henüz gerçekleşmeyen satış işlemlerine rağmen elde edilen varsayımlar doğrultusunda sağlıklı ciro simülasyonu yapabilmesi sağlanmıştır. Bu da her firma için önem arzeden nakit akışının önceden planlanmasını sağlayacaktır. Müşteri bazlı alışverişe imkan sağlayan firmalar ve web üzerinden satış yapan firmaların sattıkları ürünlerin hangi müşteri kitlesinin ilgisini çekeceğini, ürün satışa sunulmadan müşteriler tarafından ne kadar ilgi göreceği ve belirli müşteri kitlesinin ürün tercihinin neler olduğunun belirlenebilmesi firmanın müşteri ile arasında daha güçlü bağlar kurmasını sağlayacaktır.

İlk aşamada sağlıklı sonuçlar elde edilebilmesi için müşteri ve bu müşterilere ait satış verileri üzerine bazı filtrelemeler yapılarak ihtiyacımız olan veriler seçilmiştir. Bu filtreleme işlemi için doğru ve hızlı bir şekilde verilerin hazırlanabilmesi için bir uygulama yapılmıştır. Bu sayede veri karmaşasından kurtulmuş olundu. Analiz işlemine birbirine satış ve müşteri profili yakın olabilecek iki marketin verileri ile başlanmıştır. Burada amaç faklı lokasyonlardaki müşterilerin alışveriş tercihlerinin kıyaslanmasıdır. Aynı zaman da bu marketlerdeki müşteri tercihlerinin genel müşteri tercihleri ile kıyaslanması yapılmıştır. Genel müşteri tercihleri için bütün şubelere ait satış ve müşteri verileri değerlendirilmiştir ve karşılaştırma işlemi için sürece dâhil edilmiştir. Birliktelik kuralları hesap edilirken güven değeri %80 ve üstü kurallar değerlendirmeye alınmıştır. Destek değeri bize oluşturulan birliktelik kuralının ne kadar yoğun olarak tercih edildiğini gösterdiğinden, destek değeri yüksek olan birliktelik kuralları analiz sürecinde tercih edilmiştir.

Her iki markete ait birliktelik kuralları incelendiğinde bisküvi, kek ve gofret ürün gruplarına ait birliktelik kurallarının bu iki marketten alışveriş yapan müşteriler arasında rağbet gördüğü tespit edilmiştir. Bu durum üzerine her iki marketten alışveriş yapan müşteriler gruplandırılarak bu ürün gruplarını tercih eden müşteri profilleri çıkarılmıştır. Yapılan dashboard uygulaması ile müşterilere ait verilerin görsel olarak sunulması

sağlanmıştır. Müşterilere ait verilerin incelenmesine yönelik açıklamalar önceki bölümlerde açıklanmıştır. Elde edilen faydalara bakıldığında bu ürün grubunu tercih eden müşterilere ait birçok farklı veri elde edilmiştir. Çalışmamızda belirli ürünleri tercih eden müşterilere yönelik bilgiler tespit edildiği gibi tam tersi olarak belirli bir müşteri kitlesi içinde birliktelik kuralları oluşturulmuştur. Firma bu bilgiler ışığında hem kampanyaya dâhil etmek istediği ürünleri doğru müşterilere ulaştıracaktır hem de müşterilerine yönelik nokta atışı yaparak onlara ihtiyacı olan ürünleri sunarak zengin içerikli kampanyalar hazırlayabilecektir.

Kampanyanın uygulanacağı hedef kitle belirlenirken sadece birliktelik kuralına birebir uyan müşteriler değil, bu müşterilerle benzer alışverişleri yapan müşterilere de yönlendirilerek kampanyanın etkisini pozitif yönde etkileyebilir. Aynı zamanda bu yapılacak öneriler çalışmanın temel hedefleri arasındadır. Çalışmamızda yoğun bir şekilde kullandığımız Gofret, Kek -> Bisküvi kuralını göz önüne aldığımız zaman bu kurala özgü oluşturulacak bir kampanya, sadece sepetinde gofret, kek ve bisküvi ürün gruplarından en az bir ürün tercih eden müşterilere değil, gofret ve kek ürün gruplarından en az bir ürün aldıktan sonra bisküvi ürün grubundan herhangi bir ürün tercih etmeyen müşterileri de kapsamaktadır. Çünkü %80 gibi yüksek bir güven değerine sahip olan bu kuralı oluşturan kitleye sunulacak kampanya, geri kalan % 20'lik kesimi oluşturan müşterilerinde ilgisini çekebilir. Bu iki müşteri kitlesine de önerilecek bu ürün gruplarından oluşturulmuş kampanyaların ilgisini çekeceği ve satışları artıracağı düşünülmektedir.

Müşterilere özel yapılacak birkaç örnek kampanya çalışması aşağıda belirtilmiştir;

- Bunlardan birisi olan yaş bilgisine baktığımızda firma bu bilgileri inceleyip belirli bir yaş grubuna yönelik kampanya yapabilir. Örneğin 65 yaş ve üstü müşterilerine özgü bir kampanya yaparak emekli olan müşterilerine özel bir çalışma yapabilir.
- Yaş ve eğitim bilgilerini birlikte kullanarak; 25 yaş altı üniversite mezunlarına yönelik bir kampanya ile genç veya yeni mezunlara yönelik bir çalışma yapabilir.
- Meslek bilgilerini kullanarak öğretmen günlerinde öğretmenler için kampanyalar oluşturabilir.
- Belirli bir ürün grubu için kampanya oluşturmadan önce ürün gruplarının tercih edildiği zaman dilimlerini inceleyerek bu zaman diliminde alışveriş yapan müşteri gruplarına özel kampanya oluşturabilir.
- Son bir hafta, iki hafta gibi alışveriş işlemlerinde süreklilik sağlayan müşterilere ait bilgiler analiz edilerek bu müşteri kitlesi için kampanyalar sunulabilir.

Gofret, Kek -> Bisküvi birliktelik kuralının bütün şubelerdeki etkisini öğrenmek için tüm şubelere ait destek ve güven değerleri hesaplanmıştır. Bu sayede bu birliktelik kuralına yönelik yapılacak herhangi bir promosyon veya kampanyanın hangi şubelerde daha etkili ve verimli olacağı hakkında fikir edinilmiştir.

Farklı şubelerdeki güçlü birliktelik kurallarının tespiti ile farklı lokasyonda alışveriş yapan müşterilerin belirgin olan tercihleri çıkarılmıştır. Ayrıca oluşturulacak kampanyaların satış miktarlarının önceden tahmin edilmesi ile daha verimli satış kampanyaları oluşturulmasının önü açılmıştır. Müşterilerin önceki dönemlerde satın almış oldukları ürünlere ait bilgiler kullanılarak yapay sinir ağları teknikleri ile tahmin işlemleri gerçekleştirilmiştir. Bu sayede kampanya henüz fikir aşamasında iken müşteriler tarafından ne orada tercih edileceğine dair başarılı tahmin işlemi gerçekleştirilmiştir. Bu sayede birliktelik kuralları ile çok sayıda kampanyalı ürün grubu oluşturulabilecektir ve başarılı tahmin sonuçları sayesinde en verimli olacak kampanyanın müşterilere sunulması sağlanacaktır.

Bu çalışma gerçek veriler üzerinden yapılmıştır ve elde edilen sonuçlar uygulanabilir formlara dönüştürülmüştür. Mevcut perakende sektöründe kullanılabilir bir yazılımın geliştirilmesi ile verilerin daha kolay ve hızlı bir şekilde yorumlanması sağlanmıştır.

8. KAYNAKLAR

[1] A. OĞUZLAR, *Veri Ön işleme*, Erciyes Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi Sayı:21,Kayseri, Temmuz-Aralık 2003

[2] A. BAYKAL, *Veri Madenciliği Uygulama Alanları*, D.Ü.Ziya Gökalp Eğitim Fakültesi Dergisi 7,Diyarbakır, 2006

[3] Y. ÖZKAN, *Veri Madenciliği Yöntemleri*, PAPATYA YAYINLARI

[4] G.Gükay EMEL ve Ç. Taşkın, *Pazarlama stratejileri oluşturulmasında bir karar destek aracı: Birliktelik Kuralı Madenciliği* , D.Eylül Üniversitesi Sosyal Bilimler Enstitüsü Dergisi Cilt:7 Sayı:3 İzmir, 2005

[5] S.SAVAŞ, N. TOPALOĞLU ve M. YILMAZ, *Veri Madenciliği ve Türkiye'deki Uygulama Örnekleri*, İstanbul Ticaret Üniversitesi Fen Bilimleri Dergisi Yıl:11 Sayı:21 İstanbul, Bahar 2012

[6] M. TİMOR, A. EZERÇE ve U. Tuğba GÜSOY, *Müşteri profili ve alışveriş davranışlarını belirlemede kümeleme ve birliktelik kuralları analizi: Perakende Sektöründe Bir Uygulama*, İstanbul Üniversitesi İşletme Fakültesi Dergisi Yıl:22, Sayı 66, İstanbul,Şubat 2011

[7] S. ERPOLAT, *Otomobil yetkili servislerinde birliktelik kurallarının belirlenmesinde Apriori ve Fp-Growth algoritmalarının karşılaştırılması*, Anadolu Üniversitesi Sosyal Bilimler Dergisi, Eskişehir, 2012

[8] Ertuğrul Ergün, *Ürün kategori arasındaki satış ilişkisinin birliktelik kuralları ve kümeleme analizi ile belirlenmesi ve perakende sektöründe bir*, Doktora Tezi, Afyon Kocatepe Üniversitesi Türkiye, 2008.

[9] Ali Cenk GÜLCE, *Veri madenciliği apriori algoritması ve apriori algoritmasının farklı veri kümelerinde uygulanması*, Yüksek Lisans Tezi, Trakya Üniversitesi Türkiye, 2010

[10] B. Doğan, B. Erol ve A. Buldu, *Sigortacılık sektöründe müşteri ilişkileri yönetimi için birliktelik kuralı kullanılması*, Marmara Fen Bilimleri Dergisi, 2014

[11] Ö. Ç. Gülen ve S. Özdemir*, Veri madenciliği teknikleri ile üstün yetenekli öğrencilerin ilgi alanlarının analizi*, Üstün Yetenekli Eğitimli Araştırma Dergisi, 2013

[12] G. Acun, T.T. Bilgin, *Yazılım hata logları kullanılarak veri madenciliği uygulaması gerçekleştirilmesi*, Marmara Fen Bilimleri Dergisi,İstanbul, 2015

[13] M. Koç ve M. Karabatak, *Sosyal ağların öğrenciler üzerindeki etkisinin veri madenciliği kullanılarak incelenmesi*, NWSA-EDUCATION SCIENCES cilt:7 sayı:1, 2012

[14] A.KALIKOV, *Veri madenciliği ve bir e-ticaret uygulaması*, Yüksek Lisans Tezi, Gazi Üniversitesi Türkiye, 2006

[15] Eyüp Sıramkaya, *Veri madenciliğinde bulanık mantık uygulaması*, Yüksek Lisans Tezi, Selçuk Üniversitesi Türkiye, 2005

[16] Ö. DEMİR ve M. ACAR, *Sosyal bilimler sözlüğü*, Ankara Kasım 1991

[17] E.ALPAYDIN, *Zeki veri madenciliği: Ham veriden altın bilgiye ulaşma yöntemleri*, Bilişim 2000 Eğitim Semineri, İstanbul, Ocak 2007

[18] Alp S. Özdemir S ve A Kilitçi, *Veri tabanı yönetim sistemleri*, Türkmen Kitabevi, 2011

[19] G. SİLAHTAROĞLU, *Veri madenciliği kavram ve algoritmaları*, 2.Basım Papatya Y.

[20] S. Sarawagi, *Indexing OLAP Data*, IBM Almaden Research Center

[21] S. Akbulut, *Veri madenciliği teknikleri ile bir kozmetik markasının ayrılan müşteri analizi ve müşteri segmentasyonu*, Yüksek Lisans Tezi, Gazi Üniversitesi Türkiye, 2006

[22] Ceran Gülşad, *Esnek akış tipi çizelgeleme problemlerinin veri madenciliği ve genetik algoritma kullanılarak çözülmesi*, Yüksek Lisans Tezi, Selçuk Üniversitesi Türkiye, 2006

[23] P. Jacobs, *Data mining: what general managers need to know*, Harvard Management Update, 4 1999

[24] Kumar ve Wahidabanu, *Data mining association rules for making knowledgeable decisions*, Newyork IGI Global 2008

[25] HAN J. and M. Kamber, *Data mining concepts and techniques,* Morgan Kaufmann Publishers*,* USA 2001

[26] H. Akpınar, *Veri tabanlarında bilgi keşfi ve veri madenciliği*, İstanbul Üniversitesi İşletmen Fakültesi Dergisi Cilt 29-1, 2000

[27] Kıyas KAYAALP, *Asenkron motorlarda veri madenciliği ile hata tespiti*, Yüksek Lisans Tezi, Süleyman Demirel Üniversitesi Türkiye, 2007

[28] N. Duru ve M. Canbay, *Veri madenciliği ile deprem verilerinin analizi*, International Symposium Kocaeli, 2007

[29] K. Yomi, *A defect prediction method for software versioning*, Yüksek Lisans Tezi, Boğaziçi University Türkiye, 2006

[30] Y. Kılınç, *Mining association rules for quality related data in an electronics company*, Middle East Technical University Turkey, 2009

[31] Ş. Doğan ve İ. Türkoğlu, *Iron-deficiency anemia detection from hematology parameters by using decision trees*, International Journal of Science & Technology, Cilt 3, No:1, 2008

[32] A.S. Albayrak, ve Ş.K. Yılmaz, *Veri madenciliği: Karar ağacı algoritmaları ve İMKB verileri üzerine bir uygulama*, S.D.Ü. İktisadi ve İdari Bilimler Fakültesi Dergisi, Cilt 14, No: 1, 2009

[33] H.A. Ata ve İ.H. Seyrek, *The use of data mining techniques in detecting fraudulent financial statements: An application on manufacturing firms,* Süleyman Demirel Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi, Cilt 14, No 2, Isparta, 2009

[34] A.S. Bozkır, E. Sezer ve B. Gök, *Öğrenci seçme sınavında (ÖSS) Öğrenci başarımını etkileyen faktörlerin veri madenciliği yöntemleriyle tespiti*, 5. Uluslararası İleri Teknolojiler Sempozyumu (IATS'09) Karabük Üniversitesi , 2009

[35] F. Gürbüz, L. Özbakır ve H. Yapıcı, *Türkiye'de bir havayolu işletmesine ait parça söküm raporlarına ilişkin veri madenciliği uygulaması*, Gazi Üniversitesi Mimarlık Mühendislik Fakültesi Dergisi, Cilt 24, No 1, ANKARA, 2009

[36] R.J. Roiger And M.W. Geatz, *Data mining a tutorial-based primer*, Addison Wesley USA, 2003

[37] S. ÖZEKES, *Veri Madenciliği Modelleri ve Uygulama Alanları*, İstanbul Ticaret Üniversitesi Dergisi, 2003

[38] G. Karypis, E.H. Han and V. Kumar, *Hierarchical clustering using dynamic modeling*, IEEE Computer, 1999

[39] Muhsin Özgür Dolgun, *Büyük alışveriş merkezleri için veri madenciliği uygulamaları*, Yüksek Lisans Tezi, Hacettepe Üniversitesi, Ankara, 2006

[40] F. Öztemiz ve S.E. Hamamcı, *R programlama dili ve apriori algoritması kullanılarak market sepet analizi*, IDAP Sempozyumu, Malatya, 2016

[41] IBM SPSS Modeler, http://www-03.ibm.com/software/products/tr/spss-modeler (Erişim:2017)

[42] MATLAB, https://www.mathworks.com (Erişim: Haziran 2017)

[43] KEEL, http://www.keel.es/ , (Erişim Tarihi: Haziran 2017)

[44] T.T. Bilgin, *Veri akışı diyagramları tabanlı veri madenciliği araçları ve yazılım geliştirme ortamları*, Akademik Bilişim'09 - XI. Akademik Bilişim Konferansı Bildirileri, Şanlıurfa, 2009

[45] RAPIDMINER, http://rapidminer.com/, (Erişim Tarihi: Haziran 2017).

[46] WEKA, http://www.cs.waikato.ac.nz/ml/weka/, (Erişim Tarihi: Haziran 2017).

[47] M. KAYA ve S. ÖZEL, *Açık kaynak kodlu veri madenciliği yazılımlarının karşılaştırılması*, Akademik Bilişim'14 XVI Akademik Bilişim Konferansı Bildirileri, Mersin Üniversitesi, Şubat 2014

[48] R - https://www.r-project.org/ (Erişim Tarihi: Haziran 2017).

[49] SAS Enterprise Miner, https://www.sas.com/ (Erişim Tarihi: Temmuz 2017).

[50] R. Das, I. Turkoglu and A. Sengur. *Effective diagnosis of heart disease through neural networks ensembles* , Expert systems with Applications 36 (4), Elsevier, 7675-7680, 2009

[51] R. Agrawal, T. Imielinski, ve A. Swami, *Mining association rules between sets of items in large databases, In proceedings of the ACM SIGMOD*, International Conference on Management of Data (ACMSIGMOD '93) Washington USA, 1993

[52] U.T. Şimşek, *Veri madenciliği ve müşteri ilişkileri yönetiminde bir uygulama*, Doktora Tezi, İstanbul Üniversitesi, 2006

[53] R. Agrawal ve R. Srikant, *Fast algorithms for mining association rules in large databases*, Proceedings of the 20th International Conference Very Large Data Bases, San Jose, 1994

[54] F. Cemal ÖZÇAKIR, A.Yılmaz ÇAMURCU, *Birliktelik kuralı yöntemi için bir veri madenciliği yazılımı tasarımı ve uygulaması*, İstanbul Ticaret Üniversitesi Fen Bilimleri Dergisi Yıl:6 Sayı:12 Güz, 2007/2

[55] R. Agrawal, *Database Mining: A performence perspective*, IEEE Transaction on Knowledge and Data Engineering Aralık, 1993

[56] M. Houtsma and A. Swami, *Set oriented mining of association rules*, IBM Araştırma Raporu Almaden Araştırma Merkezi, California, 1993

[57] N. Kemal ERDOĞAN, B. Gülcan ve Ç. Karamaşa, *Birliktelik kuralları ve uygulamaları*, Literatür Taraması (2000-2014)

[58] R Studio , https://www.rstudio.com/ (Erişim : Haziran 2017)

[59] İleri algoritma analizi, www.ibrahimcayiroglu.com, (Erişim: Mayıs 2017).

[60] S. Yavuz ve M. Deveci, *İstatistiksel normalizasyon tekniklerinin yapay sinir ağın performansa etkisi,* Erciyes Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi Sayı:40, Kayseri, Haziran-Aralık 2012

[61] H. Aygören, H. Sarıtaş ve T. Moralı, *İMKB 100 endeksinin yapay sinir ağları ve newton nümerik arama modelleri ile tahmini,* Uluslararası Alanya İşletme Fakültesi Dergisi, Vol:4,No:1, 2012

[62] E. Öztemel, *Yapay sinir ağları*, Papatya Yayıncılık, İstanbul, 2006.

[63] F. Öztemiz ve S.E. Hamamcı, *Birliktelik kuralı ile elde edilen veri grupları için YSA ile uygulamalı tahmin gerçekleştirimi*, IDAP Sempozyum, 2017

ÖZGEÇMİŞ

Ad Soyad : Furkan ÖZTEMİZ

Doğum Yeri ve Tarihi: Malatya / Merkez 30-04-1989

Adres : Bostanbaşı mahallesi Fatih caddesi no:1 Çankaya evleri b blok daire:29

E-Posta:furkanoztemiz@gmail.com / furkanoztemiz@yandex.com

Lisans : Fırat Üniversitesi Bilgisayar Mühendisliği, 2012

Mesleki Deneyim ve Ödüller

2013 – 2017 : Esenlik Ltd. Şti - Bilgisayar Mühendisi

Yayın Listesi:

F. OZTEMIZ, S. Ethem HAMAMCI, "Market Basket Analysis Using R Programming Language and Apriori Algorithm", IDAP 2016 MALATYA (Yüksek Lisans tezinden türetilmiştir.)

F.OZTEMIZ, C. KELES, S. Ethem HAMAMCI, A.KAYGUSUZ, "The Performance Analysis of Data Mining Approach for Energy Cost Reduction in Smart Grids",ENTECH '16 ISTANBUL

F.OZTEMIZ, S. Ethem HAMAMCI, "Implementation of Applied Prediction with YSA for Data Groups with Association Rule", IDAP 2017 MALATYA (Yüksek Lisans tezinden türetilmiştir.)