

**HIZLI TÜKETİM ÜRÜNLERİ SEKTÖRÜNDE ONLINE
ALİŞVERİŞ MÜŞTERİLERİNİN SEGMENTASYONU**

**SEGMENTATION OF ONLINE SHOPPING CUSTOMERS
IN THE FAST-MOVING CONSUMER GOODS
SECTOR**

BÜŞRA KESKİN AKTAŞ

DOÇ. DR CEREN TUNCER ŞAKAR

Tez Danışmanı

Hacettepe Üniversitesi

Lisansüstü Eğitim-Öğretim ve Sınav Yönetmeliğinin

Endüstri Mühendisliği Anabilim Dalı için Öngördüğü

YÜKSEK LİSANS TEZİ olarak hazırlanmıştır.

Canım Aileme, Sevgili Eşime ve Kıymetli Dostlarıma...

ÖZET

HIZLI TÜKETİM ÜRÜNLERİ SEKTÖRÜNDE ONLINE ALIŞVERİŞ MÜŞTERİLERİNİN SEGMENTASYONU

Büşra KESKİN AKTAŞ

Yüksek Lisans, Endüstri Mühendisliği

Tez Danışmanı: Doç. Dr. Ceren TUNCER ŞAKAR

Ocak 2023, 89 sayfa

Hızlı tüketim ürünleri sektörü yıllardır var olmakla birlikte e-ticaret kapsamına son yıllarda girmiştir ve gittikçe önem kazanmaktadır. Özellikle COVID-19 pandemisinin de etkisi ile insanların alışveriş alışkanlıkları değişmeye başlamış olup, hızlı tüketim ürünleri de yoğun olarak çevrimiçi alışveriş sitelerinden satın alınmaya başlamıştır. Bu gelişmeler ile birlikte her geçen gün sektöre yeni firmalar eklenmektedir ve her şirketin pazar payını koruması gerekmektedir. Bu sebeple, nispeten yeni olan bu pazarda müşteri özelliklerinin ve davranışlarının araştırılması ile hizmetin ve kârlılığın artırılması önem kazanmaktadır. Müşteri segmentasyonu, Müşteri İlişkileri Yönetimi alanında önemli bir rol oynamaktadır. Bu tez çalışmasında, hızlı tüketim ürünleri sektöründe çok sayıda firma ile iş birliği halinde çalışan bir çevrimiçi alışveriş sitesinin müşterilerinin satın alım verileri üzerinden müşteri segmentasyonu ele alınmıştır. Öncelikle eldeki öznitelik kümesi üzerinden hangi öznitelikler ile çalışılacağı belirlenmiştir. Bu noktada literatürde yer alan güncellik, sıklık, para öznitelikleri öncelikle alınmıştır. Öznitelik seçim yöntemleri kullanılarak kuponlu sipariş sayısı yüzdesi özniteliğinin de ek olarak kullanılmaya değer olduğu tespit edilmiş ve toplamda dört öznitelik kullanılmıştır. Seçilen bu özniteliklerle uygun küme sayısı belirlenmiştir. Daha sonra kümeleme yöntemlerinden K-Ortalamlar

Kümeleme, Birch Hiyerarşik Kümeleme ve Gauss Karma Kümeleme yöntemleriyle müşteri segmentasyonu yapılmıştır. Segmentasyonun sonuçları belirlenen performans ölçütleri ile karşılaştırılmış ve şirket yetkililerinin de inceleme ve görüşleriyle K-Ortalamalar kümeleme metodunun sonuçlarının kullanılması kararlaştırılmıştır. Son olarak da K-Ortalamalar kümeleme metoduna göre belirlenen kümeler için müşteri ilişkileri geliştirme stratejileri önerilerinde bulunulmuştur.

Anahtar Kelimeler: Hızlı tüketim ürünleri sektörü, çevrimiçi alışveriş, öznelik seçimi, müşteri segmentasyonu, kümeleme algoritmaları, müşteri ilişkileri geliştirme stratejileri

ABSTRACT

SEGMENTATION OF ONLINE SHOPPING CUSTOMERS IN THE FAST- MOVING CONSUMER GOODS SECTOR

Büşra KESKİN AKTAŞ

Master's Degree, Industrial Engineering

Supervisor: Assoc. Prof. Dr. Ceren TUNCER ŞAKAR

January 2023, 89 pages

The fast-moving consumer goods sector, which has existed for years, has entered the scope of e-commerce in recent years and become increasingly important. Especially with the effects of the COVID-19 pandemic, people's shopping habits have started to change and fast-moving consumer goods have started to be purchased from online shopping sites intensively. These developments mean that new companies are added to the sector everyday and each company needs to maintain its market share. Therefore, in this relatively new market, investigating customer characteristics and behaviors, and improving service and profitability have become more important. Customer segmentation has an important role in the field of Customer Relationship Management. In this thesis, customer segmentation is studied with the purchasing data of the customers of an online shopping site working in cooperation with many companies in the fast-moving consumer goods sector. Firstly, it is determined which features will be considered from the existing

set of features. Primarily, the recency, frequency and money features in the literature are selected. Using feature selection methods, the order number percentage with coupons feature is also determined to be worth using, and four features are used in total. With these selected features, the appropriate number of sets is determined. Later, customer segmentation is carried out with K-Means, Birch Hierarchical and Gaussian Mixture Clustering methods. The results of the segmentation are compared with the determined performance metrics, and with the examination and opinion of the company officials, it is decided that the results of the K-means clustering method will be used. Finally, for clusters determined by the the K-means clustering method, customer relationship development strategies are suggested.

Keywords: Fast Moving Consumer Goods industry, online shopping, feature selection, customer segmentation, clustering algorithms, customer relationship development strategies

TEŐEKKÜR

Lisansüstü eğitim sürecim boyunca deęerli bilgi ve tecrübeleriyle beni yönlendiren hocam Doç. Dr. Ceren Tuncer Őakar'a sonsuz teşekkür ederim.

Ayrıca, hayatım boyunca desteklerini benden esirgemeyen, bana her zaman güç ve yoluma ışık olan ailem, eşim ve canım dostlarıma teşekkür ediyorum.

İÇİNDEKİLER

| | |
|--|------|
| ÖZET..... | i |
| ABSTRACT..... | iii |
| TEŞEKKÜR..... | v |
| İÇİNDEKİLER..... | vi |
| ŞEKİLLER DİZİNİ..... | viii |
| ÇİZELGELER DİZİNİ..... | x |
| SİMGELER VE KISALTMALAR..... | xii |
| 1. GİRİŞ..... | 1 |
| 1.1. Çalışmanın Motivasyonu ve Amacı..... | 1 |
| 1.2. Problemin Tanımı..... | 3 |
| 1.4. Tezin Organizasyonu..... | 3 |
| 2. GENEL BİLGİLER..... | 5 |
| 2.1. Hızlı Tüketim Ürünleri Sektörü..... | 5 |
| 2.2. Müşteri İlişkileri Yönetimi ve Müşteri Segmentasyonu..... | 6 |
| 2.3. Müşteri Segmentasyonunda RFM Analizi..... | 7 |
| 2.4. Literatür Araştırması..... | 8 |
| 3. VERİ MADENCİLİĞİ VE KÜMELEME ANALİZİ..... | 18 |
| 3.1. Veri Madenciliği..... | 18 |
| 3.1.1. Veri Madenciliği Tanımı ve Amacı..... | 18 |
| 3.1.2. Veri Madenciliği Uygulama Alanları..... | 19 |
| 3.1.3. Veri Madenciliği Tarihçesi..... | 20 |
| 3.1.4. Bilgi Keşfi Süreci..... | 22 |
| 3.1.5. Veri Madenciliği Yöntemleri..... | 23 |
| 3.2. Makine Öğrenmesi..... | 24 |
| 3.2.1. Makine Öğrenmesi Metotları..... | 26 |
| 3.3. Tez Çalışması Kapsamında Kullanılan Yöntemler..... | 33 |

| | |
|---|----|
| 3.3.1. Öznitelik Seçimi | 33 |
| 3.3.2. Küme Sayısının Belirlenmesi | 35 |
| 3.3.3. Kümeleme Analizi | 36 |
| 3.3.4. Kümeleme Performans Ölçütleri | 38 |
| 4. UYGULAMA | 40 |
| 4.1. Metodoloji..... | 40 |
| 4.2. Veri Analizi ve Verinin Kümelemeye Hazırlanması | 40 |
| 4.2.1. Genel Veri Tanıtımı | 40 |
| 4.2.2. Nümerik Özniteliklerin İstatiksel Verileri | 42 |
| 4.2.3. Sipariş Verilerine Göre Analiz | 43 |
| 4.2.4. Verinin Kümelemeye Uygun Hale Getirilmesi | 45 |
| 4.3. Öznitelik Seçimi | 46 |
| 4.3.1. Calinski-Harabasz İndeksi'ne göre Öznitelik Seçimi..... | 47 |
| 4.3.2. Davies-Bouldin İndeksi'ne Göre Öznitelik Seçimi | 49 |
| 4.4. Küme Sayısının Belirlenmesi | 50 |
| 4.5. Kümeleme Analizi | 51 |
| 4.5.1. K-Ortalamalar Kümeleme..... | 52 |
| 4.5.2. Birch Hiyerarşik Kümeleme | 58 |
| 4.5.3. Gauss Karma Model Kümeleme..... | 64 |
| 4.6. Metotların Sonuçlarının Karşılaştırması | 71 |
| 4.7. Segmentlere Göre Müşteri İlişkileri Geliştirme Önerileri | 73 |
| 5. SONUÇ VE TARTIŞMA | 78 |
| 6. KAYNAKLAR | 81 |
| EKLER..... | 88 |
| EK 1- Tez Çalışması Orjinallik Raporu..... | 88 |
| ÖZGEÇMİŞ | 89 |

ŞEKİLLER DİZİNİ

| | |
|---|----|
| Şekil 3.1. Bilgi Keşfi Süreci [40]. | 23 |
| Şekil 3.2. Denetimli Makine Öğrenmesi Süreci [45]. | 27 |
| Şekil 3.3. Denetimsiz Makine Öğrenmesi Süreci [45]. | 30 |
| Şekil 4.1. İlk Sipariş Tarihinin Aylara Göre Dağılımı | 43 |
| Şekil 4.2. İlk Sipariş Tarihinin Yıllara Göre Dağılımı | 44 |
| Şekil 4.3. Elbow Metodu Grafiği | 51 |
| Şekil 4.4. K-Ortalamalar Kümeleme, Kullanıcı Sayılarının Kümelere Göre Dağılımları | 56 |
| Şekil 4.5. K-Ortalamalar Kümeleme, Kümelere Göre Ortalama Güncellik Değerleri | 56 |
| Şekil 4.6. K-Ortalamalar Kümeleme, Kümelere Göre Ortalama Sıklık Değerleri | 57 |
| Şekil 4.7. K-Ortalamalar Kümeleme, Kümelere Göre Ortalama Para Değerleri | 57 |
| Şekil 4.8. K-Ortalamalar Kümeleme, Kümelere Göre Ortalama Kuponlu Sipariş Sayısı Yüzdesi Değerleri | 58 |
| Şekil 4.9. Birch Hiyerarşik Kümeleme, Kullanıcı Sayılarının Kümelere Göre Dağılımları | 62 |
| Şekil 4.10. Birch Hiyerarşik Kümeleme, Kümelere Göre Ortalama Güncellik Değerleri | 62 |
| Şekil 4.11. Birch Hiyerarşik Kümeleme, Kümelere Göre Ortalama Sıklık Değerleri | 63 |
| Şekil 4.12. Birch Hiyerarşik Kümeleme, Kümelere Göre Ortalama Para Değerleri | 63 |
| Şekil 4.13. Birch Hiyerarşik Kümeleme, Kümelere Göre Ortalama Kuponlu Sipariş Sayısı Yüzdesi Değerleri | 64 |
| Şekil 4.14. Gauss Karma Model Kümeleme, Kullanıcı Sayılarının Kümelere Göre Dağılımları | 69 |
| Şekil 4.15. Gauss Karma Model Kümeleme, Kümelere Göre Ortalama Güncellik Değerleri | 69 |
| Şekil 4.16. Gauss Karma Model Kümeleme, Kümelere Göre Ortalama Sıklık Değerleri | 70 |
| Şekil 4.17. Gauss Karma Model Kümeleme, Kümelere Göre Ortalama Para Değerleri | 70 |

| | |
|--|----|
| Şekil 4.18. Gauss Karma Model Kümeleme, Kümelere Göre Ortalama Kuponlu Sipariş Sayısı Yüzdesi Değerleri | 71 |
|--|----|

ÇİZELGELER DİZİNİ

| | |
|--|----|
| Çizelge 4.1. Veri Sütunları, Veri Sayıları ve Veri Tipleri..... | 41 |
| Çizelge 4.2. Nümerik Özniteliklerin İstatistiksel Veri Tablosu | 42 |
| Çizelge 4.3. Veri Temizliği Sonunda Kalan Veri Seti | 45 |
| Çizelge 4.4. Calinski-Harabasz İndeksi'ne Göre K-Ortalamalar Deneme Skorları..... | 47 |
| Çizelge 4.5. Calinski-Harabasz İndeksi'ne Göre Gauss Karma Model Deneme Skorları | 48 |
| Çizelge 4.6. Calinski-Harabasz İndeksi'ne Göre Birch Hiyerarşik Kümeleme Deneme Skorları | 48 |
| Çizelge 4.7. Davies-Bouldin İndeksi'ne Göre K-Ortalamalar Deneme Skorları ... | 49 |
| Çizelge 4.8. Davies-Bouldin İndeksi'ne Göre Gauss Karma Model Deneme Skorları .. | 49 |
| Çizelge 4.9. Davies-Bouldin İndeksi'ne Göre Birch Hiyerarşik Kümeleme Deneme Skorları | 50 |
| Çizelge 4.10. K-Ortalamalar Kümeleme Algoritma Kıyaslaması..... | 52 |
| Çizelge 4.11. K-Ortalamalar Kümeleme, 1. Kümenin Özellikleri..... | 53 |
| Çizelge 4.12. K-Ortalamalar Kümeleme, 2. Kümenin Özellikleri..... | 53 |
| Çizelge 4.13. K-Ortalamalar Kümeleme, 3. Kümenin Özellikleri..... | 54 |
| Çizelge 4.14. K-Ortalamalar Kümeleme, 4. Kümenin Özellikleri..... | 54 |
| Çizelge 4.15. K-Ortalamalar Kümeleme, 5. Kümenin Özellikleri..... | 55 |
| Çizelge 4.16. Birch Hiyerarşik Kümeleme, 1. Kümenin Özellikleri | 59 |
| Çizelge 4.17. Birch Hiyerarşik Kümeleme, 2. Kümenin Özellikleri | 59 |
| Çizelge 4.18. Birch Hiyerarşik Kümeleme, 3. Kümenin Özellikleri | 60 |
| Çizelge 4.19. Birch Hiyerarşik Kümeleme, 4. Kümenin Özellikleri | 60 |
| Çizelge 4.20. Birch Hiyerarşik Kümeleme, 5. Kümenin Özellikleri | 61 |
| Çizelge 4.21. Gauss Karma Model Kümeleme Algoritma Kıyaslaması | 65 |
| Çizelge 4.22. Gauss Karma Model Kümeleme, 1. Kümenin Özellikleri | 66 |
| Çizelge 4.23. Gauss Karma Model Kümeleme, 2. Kümenin Özellikleri | 66 |
| Çizelge 4.24. Gauss Karma Model Kümeleme, 3. Kümenin Özellikleri | 67 |
| Çizelge 4.25. Gauss Karma Model Kümeleme, 4. Kümenin Özellikleri | 67 |
| Çizelge 4.26. Gauss Karma Model Kümeleme, 5. Kümenin Özellikleri | 68 |

| | |
|--|----|
| Çizelge 4.27. Kümeleme Modellerinin Performans Kriterlerine Göre Kıyaslanması | 72 |
| Çizelge 4.28. K-Ortalamalar ve Gauss Karma Model Kümeleme Sonuçlarının Kümelerin Özelliklerine Göre Kıyaslanması | 73 |

SİMGELER VE KISALTMALAR

Simgeler

| | |
|----------------|---|
| ch | Calinski-Harabasz İndeksi |
| k | Küme sayısı |
| N | Verideki nokta sayısı |
| $tr(B_k)$ | Küme içi uzaklık karelerinin toplamı |
| $tr(W_k)$ | Kümelere arası uzaklık karelerinin toplamı |
| DB | Davies-Bouldin İndeksi |
| R_{ij} | Küme indeksi |
| d_{ij} | i ve j kümelerinin merkezleri arasındaki uzaklık |
| s_i ve s_j | Küme elemanlarının, bulunduğu kümenin merkezlerine olan ortalama mesafesi |

Kısaltmalar

| | |
|------|-----------------------------|
| FMCG | Fast Moving Consumer Goods |
| MİY | Müşteri İlişkileri Yönetimi |

1. GİRİŞ

Bu bölümde çalışmanın motivasyonu, problemin tanımı, araştırmanın amacı ve tezin organizasyonundan bahsedilmiştir.

1.1. Çalışmanın Motivasyonu ve Amacı

Günümüzün önde gelen sektörlerinden Hızlı Tüketim Ürünleri (Fast Moving Consumer Goods-FMCG) sektörü hızlı satılan, raf ömrü kısa olan ve kısa sürede tüketilen ürünlerin satıldığı ve talep gördüğü bir sektördür. Hızlı tüketim ürünlerinin satıldığı mağazalarda kullanıcılar genellikle çok vakit kaybetmekte ve istemedikleri ürünleri görmek ve araştırmak zorunda kalmaktadırlar. Bu noktada bu ürünlerin çevrimiçi alışveriş platformlarından satın alınması son yılların en önemli çalışma alanlarının oluşmasına sebep olmuştur. Çünkü çevrimiçi ortamda müşterilere ürün satmak, fiziksel mağazalara göre daha kolay ve daha az maliyetlidir. Ayrıca kullanıcıların zamandan tasarruf etmelerini sağlamaktadır.

Bununla birlikte COVID-19 salgınının da etkisiyle hızlı tüketim ürünlerinin çevrimiçi platformlardan satın alım oranları dünya genelinde çok artmıştır. Avrupa’da pandemi öncesi yüzde 9 seviyelerinde olan çevrimiçi hızlı tüketim ürünleri satın alımları, Covid-19 döneminde yüzde 65, “yeni normal” dediğimiz dönemlerde yüzde 46 seviyelerine gelmiştir. Buna ek olarak hızlı tüketim ürünlerinde e-ticaretin payı, pandemi öncesi döneme göre Meksika’da 9 puan artışla yüzde 12’ye, Birleşik Krallık’ta 7 puan artışla yüzde 13’e, ABD’de 6 puan artışla yüzde 15’e, Kore’de 4 puan artışla yüzde 22’ye, Çin’de 5 puan artışla yüzde 30’a yükselmiştir. [1]

Türkiye’de de bu sektörün önemi giderek artmaktadır. Günden güne rakip sayısı çoğalmaktadır ve sektöre katılan her yeni rakiple şirketlerin var olan pazar payları küçülme riski altında kalmaktadır. Bu durum, şirketleri tüketicilerin alışveriş alışkanlıklarındaki büyük değişimi anlamaya çalışmaya itmektedir. Bu değişimi anlamamanın en önemli adımı da müşterileri segmentlere ayırarak yönetmektir.

Müşteri segmentasyonu, yaygın olarak Müşteri İlişkileri Yönetimi (MİY) kapsamında yer almaktadır. MİY, firmaların mevcut ve potansiyel müşterileri ile kurdukları ilişkiyi en verimli ve en kârlı hale getirmek için kullandıkları yöntemlerin toplamıdır ve özellikle hizmet sektöründe çok önemlidir. MİY en basit anlatımıyla dört adımdan oluşmaktadır. Bunlar; müşteri tanımlama, müşteri çekme, müşteriye elde tutma ve müşteri geliştirmedir. Bu aşamalardan özellikle ilk adım çok kritiktir ve burada müşterilerin segmentasyonu söz konusudur. Bu tez çalışmasında da MİY'in müşteriye tanıma adımında yapılan segmentasyon çalışması yürütülecektir.

Yapılan literatür çalışmalarında özellikle çevrimiçi hızlı tüketim ürünleri satın alımları için detaylı segmentasyon çalışmalarına rastlanmamıştır. Bu yüzden bu çalışmada, 2015 yılından beri çevrimiçi hızlı tüketim ürünleri sektöründe yer alan ve pandeminin de etkisiyle satış oranları artış gösteren bir şirketin müşteri davranış verileri kullanılarak müşteri segmentasyonu yapılmasına ve belirlenen segmentlerin özelliklerine göre pazarlama birimine bilgi sağlanmasına odaklanılmıştır.

Tezin genel amacı, hızlı tüketim ürünleri sektöründe online alışverişte müşteri hizmeti politikasının ve kârlılığın iyileştirilmesi için müşteri segmentasyonuna dayalı bir yaklaşım geliştirilmesidir.

Bu tez öncelikle çevrimiçi hızlı tüketim ürünleri sektörü müşterilerinin segmentasyonunda kullanılacak uygun öznitelikleri ortaya çıkarmayı amaçlamaktadır. Elimizde olan güncellik (Recency-R), sıklık (Frequency-F), para (Monetary-M), kupon kullanım sayısı yüzdesi, kupon kullanım tutarı yüzdesi, kampanya kullanım sayısı yüzdesi, kampanya kullanım tutarı yüzdesi ve gün farkı gibi kullanıcı davranış verilerinden hangilerinin kullanımının iyi bir segmentasyon oluşturulmasına yarayacağı tespit edilmek istenmektedir. Sonrasında bu özniteliklerle en iyi şekilde çalışacak kümeleme algoritması bulunmaya çalışılacaktır. Son olarak elde edilen kümeler için farklı müşteri ilişkisi stratejileri önerilecektir. Yapılacak çalışmalar şu anda bu sektörde hızla büyümekte olan gerçek bir firmanın verileri ile yapılacaktır.

1.2. Problemin Tanımı

Hızlı tüketim ürünleri eski bir sektör olmasına rağmen çevrimiçi alışveriş kapsamına girmesi görece yeni bir gelişmedir. Bu sektöre yeni giriş yapan birçok firma olduğu gibi uzun yıllardır e-ticaret yapan şirketler de hızlı tüketim ürünleri tarafında ilerlemeye çalışmaktadır. Özellikle pandeminin de etkisiyle sektör giderek büyümekte ve rekabet her geçen gün artmaktadır. Literatür bölümünde de görülebileceği gibi diğer sektörlerde birçok müşteri segmentasyonu çalışması olmakla birlikte bu sektörde detaylı bir çalışma bulunamamıştır. İncelenen çalışmalarda bu sektör için güncellik, sıklık ve para özniteliklerine ek olabilecek başka öznitelikler üzerinde durulmadığı gözlemlenmiştir.

Bu tez, hızlı tüketim ürünleri sektöründe çevrimiçi satış yapan firmaların müşteri ilişkileri yönetimini için kullanabilecekleri bir çalışma olacaktır. Geliştirilen yöntem gerçek bir firmanın müşteri davranışları verileriyle uygulanmıştır ve bu sayede sonuçlar aynı sektördeki diğer firmalara yol gösterici olacaktır. Ayrıca farklı makine öğrenmesi yöntemleri ile yapılan kümelemelerin sonuçları detaylı olarak sunulacaktır. Bu açıardan çalışmanın hem pratik hem de teorik katkılarının olacağı düşünülmektedir.

1.4. Tezin Organizasyonu

Toplamda beş bölümden oluşan bu tez çalışmasında ilk bölüm Giriş bölümüdür ve bu bölümde çalışmanın motivasyonundan, problemin tanımından ve araştırmanın amacından bahsedilmiştir.

İkinci bölüm genel bilgilerden oluşmaktadır. Bu bölümde tez çalışması kapsamında müşteri satın alım verileri kullanılan şirketin sektör bilgilerine, müşteri ilişkileri yönetimi ve müşteri ilişkileri yönetiminde müşteri segmentasyonunun önemine, müşteri segmentasyonunda RFM özniteliklerinin kullanımına ve bunlara göre yapılmış literatür araştırması bilgilerine yer verilmiştir.

Üçüncü bölümde veri madenciliği tanımlarından, kullanım amaçlarından, uygulama alanlarından, tarihçesinden, bilgi keşfi sürecinden ve veri madenciliği yöntemlerinden bahsedilmiştir. Yine aynı bölümde makine öğrenmesi tanım ve bilgilendirilmelerine, makine öğrenmesinde kullanılan metotlara yer verilmiştir. Son olarak da tez çalışması kapsamında özniteliklerin seçiminde kullanılan kriterlerden, küme sayısının belirlenmesi için kullanılan metotlardan, kullanılacak kümeleme metotlarından ve kümeleme performanslarının kıyaslanabilmesi için kullanılan performans ölçütlerinden bahsedilmiştir.

Dördüncü bölüm uygulama bölümüdür ve bu bölümde ilk olarak metodolojiden bahsedilmiş. Sonrasında eldeki müşteri verilerinin kümelemeye uygun hale getirilebilmesi için yapılan analizlerden ve ön işleme işlemlerinden bahsedilmiştir. Daha sonra sırasıyla öznitelik seçimi, küme sayısının belirlenmesi, kümeleme algoritmalarının uygulanması ve sonuçlarının paylaşılması adımları gerçekleştirilmiştir. Analizler Python kütüphaneleri kullanılarak gerçekleştirilmiştir. Son olarak da yapılan analizlerin sonuçları farklı metotlara göre kıyaslanarak, kullanılmasına karar verilen modelin sonuçlarına göre ilgili segmentler için müşteri ilişkileri geliştirme önerilerinde bulunulmuştur.

Beşinci bölüm sonuç ve tartışma bölümüdür ve bu bölümde veri madenciliğinin segmentasyon çalışmaları ve pazarlama çalışmaları için öneminden bahsedilmiştir. Tez çalışması kapsamında yapılan çalışmaların özetinden ve sonuçlarından bahsedilmiştir. Son olarak da segmentasyon çalışmalarının belli aralıklarla tekrarlanan çalışmalar olması gerektiğinden bahsedilmiştir.

2. GENEL BİLGİLER

Bu bölümde, tezdeki asıl çalışma öncesinde bilgi sahibi olunması gereken genel bilgiler aktarılmıştır. Bunlar, çalışmanın yapıldığı sektör bilgileri, yapılan çalışmanın hangi kapsamda olduğu, kullanılan analiz yöntemi ve literatür çalışmasıdır.

2.1. Hızlı Tüketim Ürünleri Sektörü

Tüketim, üretilmiş olan mal ve hizmetlerin son kullanıcı tarafından ihtiyaç veya isteklerinin giderilmesi amacıyla kullanılmasıdır. Bu noktada işletmelerin bir amacı da işletme tarafından üretilmiş ürünlerin son kullanıcı tarafından tüketilmesini sağlamak ve söz konusu tüketim ürünlerini devamlı olarak üretmeye devam etmektir [2].

Seri üretim ve kitlesel tüketimin gelişmesine bağlı olarak ortaya çıkan hızlı tüketim ürünleri, ihtiyaç ve istekleri gidermek için kimi zaman zorunlu olarak dahi tüketilen ürünlerin tümüdür. Yani hızlı tüketim ürünleri, tüketiciler tarafından sürekli kullanılan, hızlı tüketilen, talep düzeyi yüksek ve maliyetleri düşük olan ürünlerdir. Hızlı tüketildikleri ve müşteriler tarafından talep düzeyleri yüksek oldukları için raf ömürleri kısadır ve hızlı hareket eden ürünler olarak kabul edilirler.

Hızlı tüketim ürünlerinin özelliklerine bakıldığında; tüketiciler tarafından satın alma aşamasında üzerine çok düşünülmeyen, sürekli alınan ve bu süreklilik içerisinde işletmeler tarafından büyük miktarlarda üretilip piyasaya sürülen bu ürünler, hızlı bir şekilde tüketildikleri için rafta durma süreleri ve stok süreleri oldukça kısa olan ürünlerdir. Hızlı tüketim ürünlerine örnek olarak; şarküteri, içecek, paketli veya paketsiz gıdalar, kişisel bakım ve kozmetik, reçetesiz sağlık ürünleri, temizlik malzemeleri, kırtasiye malzemeleri ve benzeri gibi geniş bir ürün yelpazesini kapsayacak şekilde örnekler verilebilir.

Hızlı tüketim ürünleri ya da paketlenmiş tüketici ürünleri, hızlı satılan ya da kısa süre içerisinde kolaylıkla tüketilen, kullanımları gün, ay ya da yıl bazında sınırlandırılan düşük maliyetli ürünlerdir. Hızlı tüketim ürünleri sektörü ise tüketicilerin perakendeci raflarında

gördüğü hızlı tüketilen her şeyi kapsar. Tiwari'ye göre [3], hızlı tüketim ürünlerinin tüketici gözünden ana özellikleri sık satın alınan, düşük fiyatlı ve düşük sadakate sahip olması iken; firma açısından ise yüksek satış hacimli, düşük kar marjlı, geniş dağıtım ağına ve hızlı stok döngüsüne sahip olması şeklindedir.

Türkiye'de hızlı tüketim sektörünün 2022 yılının ilk 8 aylık döneminde %85 oranında büyüme kaydettiği ve yine bu dönemde e-ticaretin %74 oranında büyüdüğü raporlanmıştır [4]. Oranlara bakıldığında bu sektöre duyulan ilginin özellikle son dönemlerde arttığı söylenebilir. Bu sebepler göz önüne alınarak bu tez çalışmasında hızlı tüketim ürünleri sektöründe faaliyet gösteren bir çevrimiçi alışveriş platformunun kullanıcı verisi incelenmiştir.

2.2. Müşteri İlişkileri Yönetimi ve Müşteri Segmentasyonu

Üretici firmaların ürettiği ürün veya hizmetleri, ihtiyaç ve isteklerini karşılamak amacıyla ürüne biçilmiş fiyatı ödeyerek satın alan ve tüketen kişiye müşteri denir.

MİY, kuruluşların mevcut veya hedef müşterileri ile kurdukları ilişkiyi en etkili ve en karlı hale getirmek için kullandıkları yöntemlerin bütünüdür. MİY en basit anlatımıyla dört adımdan oluşmaktadır. Bunlar; müşteri tanımlama, müşteri çekme, müşteriyi elde tutma ve müşteri geliştirmedir. Bu aşamalardan özellikle ilk adım çok kritiktir ve burada müşterilerin segmentasyonu söz konusudur. Segmentasyon işleminin birçok türü vardır ve bir amaç olarak değil, doğru kullanıcıya dokunarak karlılığı artırma amacına giden çok önemli bir araç olarak kullanılmaktadır.

Segmentasyon işlemi bir kere yapıлып bırakılabilecek bir işlem değildir. Dinamiktir ve sürekli tekrarlanması gerekmektedir. Pazar koşulları göz önüne alınarak segmentasyon için kullanılan parametrelerin sürekli gözden geçirilmesi ve müşteri sınıflandırmasının sürekli yapılması gerekmektedir. Literatürde segmentasyon işlemi gerçekleştirmek için dikkate alınan birçok farklı kriter ve önerilen birçok teknik bulunmaktadır.

İşletmeler, ürettikleri ürün/hizmetlerin hedef pazardaki müşterilerinin hangi ihtiyaç veya isteğini karşıladığını göz önünde bulundurarak, benzer özelliklerine göre müşterilerini

sınıflandırılabilir ve bu sınıflandırmaya göre bütün pazarı daha özel niteliklere sahip alt pazarlara bölebilir. Bu şekilde alt pazarlarla özel olarak ilgilenirken hedef dışında kalan özellikteki müşterilere zaman ve para kaynaklarını harcamamış olur. Pazardaki tüm müşterileri tatmin edemeyeceği için müşterileri alt segmentlere bölerek belirlediği hedef pazarında maksimum verim almak isteyen ve bunun için etkili bir kampanya hazırlamak isteyen işletmeler, hedef pazarlarındaki müşterileri iyi tanımalı, onlar hakkında veri toplayarak bu verileri iyi analiz etmeli ve bu analizler sonucunda onları en tatmin edecek ürün veya hizmeti üretmelidir.

Müşteri segmentasyonu yapılırken genel olarak şu 4 özellik baz alınır:

- Demografik özellikler (yaş, cinsiyet, eğitim durumu, vs.)
- Coğrafi özellikler (Bölgeler, ülkeler, şehirler, vs.)
- Psikografik özellikler (Yaşam tarzı, ilgi alanları, vs.)
- Davranışsal özellikler (Tüketim ve harcama özellikleri, vs.)

Bu tez çalışmasında FMCG sektöründe 2015 yılından beri faaliyet gösteren bir firmanın davranışsal özelliklerinden oluşan gerçek verileri kullanılarak müşteri segmentasyonu yapılacaktır.

Davranışsal, yani tüketim ve harcama özelliklerinin kullanılmasının sebebi, bu sektörde kullanıcı verilerinin soyut olmasından kaynaklanmaktadır. Başka bir tabirle vatandaşlık numarası gibi belirgin bir veri olmadığı için ad, soyad, yaş, cinsiyet gibi veriler yüzde yüz doğru değildir. Hatta büyük bir çoğunluğu gerçeği yansıtmamaktadır. Bu yüzden elimizdeki en doğru veriler davranışsal verilerdir.

2.3. Müşteri Segmentasyonunda RFM Analizi

Müşterilerini tanımak ve onları tanımlayabilmek işletmeler için en kritik adımdır. İşletmelerin varlıklarını sürdürebilmeleri yani sürdürülebilir bir iş modeli geliştirebilmeleri müşterilerini analiz etmelerine, anlamalarına ve analiz sonuçlarına göre tüm süreçlerini dönüştürmelerine bağlıdır. Sohrabi ve Khanlari [5], müşterilerin

tamamının finansal anlamda işletmelere aynı düzeyde cazip olmadığını ve karlılıklarının belirlenerek sonuçlara göre uygun kaynak ayırımı yapmanın kritik olduğunu belirtmiştir.

RFM Analizi, müşterilerin bazı satın alma alışkanlıklarını analiz ederek müşteri segmentlerini belirlemek için kullanılan bir istatistiksel analiz yöntemidir. Müşteri alışverişlerinin güncellik, sıklık ve para değerlerini tespit edip bunlara göre etiketleme yapma işlemidir. Yani müşterilerin geçmiş finansal davranışlarının analiz edilerek gelecekteki davranışlarını tahminlemek ve pazar bölümlendirme yapabilmek amacıyla kullanılan bir yöntemdir [6]. Bu analiz sayesinde işletmeler müşteri yaşam döngüsünün farklı bölümlerinde yer alan müşterileri tespit ederek, bu müşteri gruplarına uygun pazarlama stratejilerini farklılaştırabilmektedirler [7].

Güncellik değeri, müşterinin son satın alım işleminden analiz zamanına kadar olan periyodun uzunluğunu temsil eder. Sıklık değeri, dikkate alınan zaman dilimi içerisinde yapılan satın alma işlemi sayısını ifade eder. Para değeri, dikkate alınan zaman dilimi içerisinde yapılan satın alma işlemleri için harcanan toplam tutarı ifade etmektedir.

2.4. Literatür Araştırması

Bu bölümde müşteri ilişkileri yönetimi, müşteri segmentasyonu ve kümeleme algoritmaları konuları ile alakalı literatür araştırması yapılmıştır.

Tao [8], yapmış olduğu çalışmada MİY'in tarihçesinden bahsetmiş daha sonra müşteri beklentilerini analiz ederek müşteri memnuniyetine etki eden faktörlere değinmiştir. Tao bu çalışmasında müşteri memnuniyetinin nasıl değerlendirileceğinin üzerinde durulmasının gerekliliğine ve uygun önlemlerin planlanmasının önemine yer vermiştir. Buradaki en önemli konuların müşterilerin gerçek ihtiyaçlarını karşılamak için "müşteri odaklı" geliştirme stratejisinin oluşturulması (MİY pazarlama stratejileri) ve müşteri memnuniyetini artırırken müşteri beklentilerinin fark edilmesi ve yürütülmesi olduğundan bahsedilmiştir. Ayrıca pazardaki rekabetin doğasının yüksek kaliteli müşteriler arasındaki rekabet olduğuna değinilmiştir.

Hjort ve diğeri [9], yaptıkları çalışmada müşterilerin satın alma ve iade davranışlarına göre müşterileri segmente edebilmeyi ve “herkese uyan tek beden” yaklaşımına müşterilerin bakış açısını bularak e-ticaret işine uygun olup olmadığını tespit etmeyi amaçlamışlardır. Bunun için müşteri işlemlerini inceleyerek satış ve iade verilerinin analizini yapmışlardır ve her müşterinin net katkı marjları ölçmüşlerdir. Çalışma sonucunda müşterilerin heterojen bir şekilde davrandığını tespit edilmiş olup satış ve iade verilerine göre yapılmış bir segmentasyonunun farklılaştırılmış bir hizmet sunum yaklaşımını kolaylaştırabileceğini ortaya koymuştur. Yani bu çalışma ile literatürün belirttiği “herkese uyan tek beden” teorisinin modasının geçtiği gösterilmiştir. Ayrıca müşteri davranışlarında önemli farklılıklar bulunuyorsa, bu farklılıkların daha ayrıntılı olarak araştırılması gerektiğini de eklemişlerdir.

Vicente ve diğeri [10], yapmış oldukları derleme makalede, 2015-2019 yılları arasında hazırlanmış 17 makale üzerinden MİY’de yaşanan gelişmelerin iş inovasyonuna etkilerini anlatmışlardır. Çalışma sonucunda, kuruluşların, performanslarında iyileştirme arayışında olmaları durumunda en alakalı faktörlerden ikisi olarak MİY ve İnovasyon ortak çalışmasını ele almaları gerektiğini ve ayrıca MİY’in, müşterinin firmaya karşı sadakatini güçlendirme ve kuruluşların değişen iş ortamına uyum sağlamak amacıyla süreç inovasyonu uygulamak için önemli bir araç olarak kullanılabileceğini belirtmişlerdir.

Karakoç ve Özkan [11], iki amaçlı bir çalışma yapmışlardır. Bunlardan ilki bir hastanenin hasta verilerini kullanarak LRFM (Length-L, Recency-R, Frequency-F, Monetary-M) analizi metodu ile karlı müşterileri belirleyebilmektir. İkincisi ise karlılık açısından daha büyük öneme sahip olan müşterilerin hastaneye bağlılık oranını ve hastanede harcadıkları para miktarını artırabilmek için uygulanabilecek pazarlama stratejilerini belirlemektir. Bu amaçlar için klasik RFM’in yanına uzunluk (length-L) boyutunu da eklemişlerdir. Bu parametre hastanın hastanede yaptığı ilk kayıt tarihi ile son geliş tarihi arasında geçen zamanı ifade etmektedir. Hangi kümeleme metodunu kullanacaklarını seçebilmek için kümeleme doğrulama analizi yapmışlar ve dört kümeli Hiyerarşik Kümeleme Metodu’nun kullanımına karar vermişlerdir. Ward metodunu ise katsayıları hesaplama metodu olarak seçmişlerdir. Kümelenen müşterileri Müşteri Değer Matrisi ve Müşteri İlişki matrisine koyarak yer aldıkları bölgeleri belirlemişlerdir. Bu veriler üzerine

hastaların LRFM deęerleri ve demografik zellikleri (kurum, yař, cinsiyet, fatura tipi ve il deęiřkenleri) dikkate alınarak ve CART metodunun Yinelemeli Ayırma (Recursive Partitioning- RPart) algoritması kullanılarak, elde edilen kmelerin oluřmasını saęlayan en nemli deęiřkenler ve ayırıtıran deęerleri bulmak amalanmıřtır. Analiz sonucunda sınıfların ortaya ıkmasında en nemli deęiřkenlerin sıklık, gncellik, uzunluk ve hastanın sigortalı olduęu kurum olduęu tespit edilmiřtir. Son olarak elde edilen veriler zerine de drt kme iin de ayrı ayrı stratejik nerilerde bulunulmuřtur.

Ahmed ve dięerleri [12], tm dnyada satıřı olan bir video oyununun satıř verilerinden elde edilen seyrek ok boyutlu verileri kullanarak Gauss Karma Kmeleme ve K-Ortalamlar Kmeleme modellerini denemiřlerdir. Veri olarak meknsal koordinatlar, satıřlar ve derecelendirme puanları alınmıřtır. Denemeler sonucu derecelendirme puanlarının iyi bir kmeleme sonucu vermedięi dřnldę iin analizden ıkarılmıřtır. 2 boyutlu bir verinin kmelenmesinde iki model de iyi sonular vermemiřtir. Bu durumun boyut sayısının az olmasından dolayı yařandığı dřnlmř ve daha iyi bir kmeleme yapabilmek iin daha ok znitelięe ihtiya olduęunu sylemiřlerdir. K-Ortalamlar kmeleme modeli Gauss Karma Kmelemeye gre daha hızlı alıřmıřtır. Kme sayısının bilindięi durumlarda daha iyi sonular elde edilebileceęini tartıřmıřlardır.

Veri madencilięi, firmaların ellerindeki verilere bakarak mřteri eęilimlerini arařtırmasına ve sonrasında bu arařtırmalara dayanarak mřteri iliřkilerini geliřtirmesine yardımcı olan gl bir tekniktir ve MİY’de nemli bir aratır. Ancak veri madencilięin bazı dezavantajları vardır. Cheng ve Chen [13] alıřmalarında, anlam kurallarını ıkarmak iin RFM karakteristiklerinin nicel deęerini ve K-Ortalamlar algoritmasını kaba kme (Rough Set-RS) teorisi ile birleřtiren yeni bir prosedr nermektedir. alıřmanın  amacı bulunmaktadır. Bunlar: RS algoritmasını geliřtirmek iin srekli karakteristikleri ayırmak, sbjektif grře dayalı olarak mřteri baęlılıęı ıktı olacak řekilde 5 kmeye ayrılacak řekilde mřteri deęerini kmelemek ve ardından doęruluk oranında hangi sınıfın en iyi olduęunu grme ve son olarak MİY’i glendirmek iin mřterinin zelliklerini ğrenebilmektir. Uygulamayı yapabilmek iin Tayvan elektronik endstrisinde pratik olarak toplanan bir C-řirketi veri seti kullanılmıřtır. İlk olarak kmeleme algoritmasına girdi olması amacıyla nicel deęer elde etmek iin RFM modelini kullanılmıřtır. Daha sonra mřterileri kmelemek iin K-Ortalamlar algoritması

kullanılmıştır. Son olarak, işletmelerin mükemmel bir MİY yürütmesine yardımcı olan sınıflandırma kurallarını çıkarmak için kaba kümeler (LEM2 algoritması) kullanılmıştır. (RS teorisinde, karar kuralları genellikle belirli bir karar tablosundan uyarılmaktadır.) Çalışma sonuçlarına göre F karakteristiği ve M karakteristiğine göre High olan müşterilerin şirketin satışlarının %89,59'unu karşıladığı görülmüştür. Bu müşteriler C-şirketi ile ticaret yapan tüm müşterilerin yaklaşık %26,18'ine eşittir. İstatistiksel sonuçlara göre de bulunan sonuç Pareto ilkesiyle örtüşmektedir. Bu müşteriler müşteri ilişkileri yönetimi için en önemli kaynaktır. Bu sonuç, çalışmanın C-şirketinin ampirik vaka çalışmasının deneysel sonuçları aracılığıyla hedef müşterilere odaklanmasına yardımcı olabileceğini kanıtlamaktadır.

Tavakoli ve diğerleri [14], yaptıkları çalışmada müşteri segmentasyonunun müşteri ilişkileri yönetiminin ilk adımı olan müşteri tanıma ve doğru stratejileri geliştirebilmek açısından öneminden bahsetmişlerdir. Müşteri segmentasyonu için R+FM adlı yeni bir RFM çalışma alanı önererek R'yi diğer iki özellikten (F ve M) ayırmışlardır. Bunu yapmalarının sebebini ise yeniliğin yalnızca son satın alma zamanını göstermesi fakat F ve M değerlerinin müşterinin sadakatini göstermesi olarak açıklamaktadırlar. Önerdikleri bu modeli, Alexa Sıralamasına göre İran'da en çok ziyaret edilen web siteleri arasında dördüncü sırada yer alan Digikala1 şirketinin gerçek verilerine dayalı olarak değerlendirilmişlerdir. İlk önce ilgili şirket yöneticilerinin talebi de dikkate alınarak segmentasyon modeli uygulanmış ve R değeri için Aktifler, Zaman Aşımına Uğrayanlar ve Süresi Dolanlar olmak üzere 3 segment elde etmişlerdir. Bunun üzerine K-Ortalamalar algoritması ile Aktifler segmenti Yüksek Değer, Orta Değer-Yüksek Parasal, Orta Değer-Yüksek Frekans ve Düşük Değer olacak şekilde 4 segmente, Zaman Aşımına Uğrayanlar ve Süresi Dolanlar segmentleri ise Yüksek Değer, Orta Değer Düşük Değer olacak şekilde 3 segmente ayrılmıştır. Son olarak ise bulunan bu segmentler için bir SMS kampanyası yürüterek önerilen modelin diğer Digikala kampanyalarına kıyasla satın alma sayısında ve gelirinde artışa yol açtığını ispat etmişlerdir.

Çalış ve Baynal [15], ülkemizdeki büyük bir bankanın birinci sınıf şubelerinden birinin verilerinden yararlanarak veri madenciliği çalışması yapmışlardır. Çalışmalarında kullanabilmek amacıyla, bankadan müşterilerin 12 farklı değişkenden oluşan kişisel özelliklerini (bireysel kredi müşterilerinin yaş, cinsiyet, medeni hal, öğrenim durumu,

aylık gelir, ev, araç, çocuk sahibi olma durumu, eş geliri, ödeme durumu, banka maaş müşterisi olma durumu ve çalışma şekli) talep etmişlerdir. Daha sonra SPSS Clementine ile bu verilere K-Ortalamlar metodu ile kümeleme analizi yapılarak müşterilerin değerlendirmesi amaçlanmıştır. Kümeleme analizi sonucunda 3 küme elde edilmiş ve bu kümelerin özelliklerinden ve müşteri profillerine göre satış stratejilerinden bahsedilmiştir.

Wei ve diğerleri [16], Tayvan'da kuaförlük sektörünün gelişimiyle birlikte fazlasıyla çalışılmak istenen bir alan olmasından dolayı ve Tayvan kuaför endüstrisinin hizmetleri etkin bir şekilde pazarlamak için karlı müşterileri belirlemeye ve kayıp müşterileri elde tutmaya ihtiyaç duymalarından dolayı yaptıkları çalışmada ünlü bir kuaförün müşterilerini segmente etmişlerdir. Kuaför salonunu etkin bir şekilde tanımlamak için Kendi Kendini Düzenleyen Haritalar (SOM) ve K-Ortalamlar yöntemleri ile RFM modelinin birleştirilmesiyle veri madenciliği teknikleri benimsenmiştir. Çalışma sonucunda SOM yöntemi ile 11 küme oluşturulmuştur. Daha sonra RFM modelinde uygulanmak üzere K-Ortalamlar tekniği kullanılmıştır. RFM modelinin analiz sonuçlarına göre, kuaför salonundaki müşteriler 'Sadık Müşteriler', 'Potansiyel Müşteriler', 'Yeni Müşteriler' ve 'Kaybedilen Müşteriler' olmak üzere 4 müşteri tipinde gruplandırılabilir. Daha sonra bu 4 müşteri tipi için pazarlama stratejileri önerilmiştir.

Zalaghi ve Varzi [17], yaptıkları çalışmada müşterilerin sadakat oranını hesaplamak amacıyla bir yöntem sunmuşlardır. Bu çalışmada ilk olarak K-Ortalamlar kümeleme yöntemi kullanılarak belirli kümelerdeki müşterileri kategorize ettiler. Daha sonra, ağırlık (W) değerleri ile RFM modelindeki her değişkenin ağırlığını temsil eden bir WRFM sürecine göre sınıflandırmaya yardımcı olan karakteristikleri tanımladılar. Bu ağırlıklar çalışmanın uygulandığı sektörle ilgili değişkenlerin önemini bilen firma uzmanları tarafından belirlenmiştir. Spearman korelasyon katsayısı, sadık ve sadakatsiz müşterileri aydınlatmayı amaçlayan tüm değişkenler için hesaplanmıştır. Çok amaçlı genetik algoritma, RFM modelini oluşturmadan önce bu özellik kümesini tanımlamak için kullanılmıştır. Bu algoritmanın amacı, özelliklerin sayısını azaltmak ve tahmin gücünü artırmak olarak tanımlanmıştır. Son aşamada, dört küme oluşturulmuş ve bağlılık oranları açısından en kötü ve en iyi müşterilere bazı stratejiler uygulanmıştır. Uygulamadan elde

edilen sonuçlara göre, önerilen yöntemin müşterilerin yaşam sürelerinin durumunu belirlemede yüksek hassasiyete sahip olduğunu göstermiştir.

Soeini ve Fathalizade [18], yaptıkları çalışmada klasik RFM modeline ilk satın alımından bu yana geçen süre ve maliyet (Cost-C) olmak üzere iki parametre daha dahil ederek genişletilmiş bir RFM modeli (LRFMC) geliştirmişlerdir. Bu çalışmanın amacını, doğrudan pazarlama için hedef seçme bilgisini keşfetmek için kapsamlı bir metodoloji sunmak olarak tanımlamışlardır. Makalede ilk önce MİY konseptinden daha sonra RFM modelinden bahsedilmiş olup sonrasında genişletilmiş model önerilmiştir. Uygulama bölümünde ise sigorta endüstrisindeki bir vakayı incelemişlerdir. Ayrıca bu bölümde Kotler tarafından tanımlanan Müşteri Yaşam Boyu Değerini de kullanmışlardır. Çalışmanın sonucunda RFM ve LRFMC kıyaslanmış olup değiştirilmiş modelin (LRFMC) güven ve destek açısından RFM modelinden daha iyi olduğu tespit edilmiştir.

Wei ve diğerleri [19], yaptıkları çalışmada uzunluk parametresini ekleyerek RFM modeli aracılığıyla genişletilmiş LRFM modelini uygulamış ve bir çocuk diş kliniğindeki hastaları segmentlere ayırmak ve önemli hastaları hedeflemek için kendi kendini organize eden haritaları kullanmıştır. Yani küme sayısını belirlemek için veri seti Kendi Kendini Düzenleyen Haritalar SOM tarafından kümelenebilir. 2258 hastanın verisini kullanmışlardır. Bu veriler her hasta için üyelik numarası, cinsiyet, doğum tarihi, ilk ziyaret tarihinden son ziyaret tarihine kadar geçen gün sayısı, son ziyaret tarihi, ziyaret sıklığından ve parasal değerden oluşmaktadır. Bu hastalar için on iki küme bulmuşlardır ve geliştirilen bu kümeler ve Müşteri İlişkileri Matrisi sonuçlarına göre diş kliniği için pazarlama stratejileri önermişlerdir. Her küme için genel ortalamadan daha büyük olan LRF değişkenlerinin değerleri tanımlamışlardır. Sonuçlar, ortalamanın üzerinde LRF değerlerine (454 hasta) sahip üç kümenin çekirdek hasta olarak görülebileceğini göstermektedir.

Sari ve diğerleri [20], yaptıkları derleme makalesinde müşteri segmentasyonuna değinmişler ve müşteri segmentasyon yöntemlerini veri işlemeye dayalı olarak sınıflandırmışlardır. Çalışmalarında müşteri segmentasyonunun ne olduğunu ve neden gerektiğini anlattıktan sonra segmentasyon için gerekli olan veriler ve yöntemlerden

bahsetmişlerdir. Segmentasyonu dahili veri ve harici veri olarak gruplamışlardır. Müşteri profilini ve satın alma geçmişini dahili veri olarak, sunucu günlüğünü, çerezler ve anket verilerini ise harici veri olarak kabul etmişlerdir. Bu verilerin İş Kuralı, Magento, Müşteri Profili Oluşturma, Niceliksel Üyelik, RFM Hücre Sınıflandırma Gruplaması, Denetimli Kümeleme, Müşteri Benzerliği Kümelemesi, Satın Alma Yakınlığı Kümelemesi ve Denetimsiz Kümeleme gibi yöntemler kullanılarak işlenebileceğinden bahsetmişlerdir. Son olarak da segmentasyon sürecine değinmişlerdir.

Özmen, Delice ve Aydoğan [21], telekomünikasyon sektörü müşterilerine uyan bir sistem tasarlayarak, sektördeki rekabet gücünü iyi bir hale getirebilmeyi amaçladıkları bu çalışmada ülkemizdeki telekomünikasyon şirketlerinden birinin müşterilerini tanımlayabilmek, anlayabilmek ve beklentilerini karşılayabilmek için müşteri segmentasyonu çalışması yapmışlardır. Başlangıçta bu segmentasyon için gereken en önemli kriterleri belirlemişlerdir. Daha sonra bu verileri Parçacık Sürü Optimizasyonu metodu ile kümelemişlerdir. Çalışmanın kapsamı veri toplama, ön işleme ve kümeleme adımlarından oluşmaktadır. Müşteri verisi olarak Müşteri, Arama ve Fatura ve Ödeme ana başlıklarının altında 10 kriter kullanılmıştır. Uygulamanın sonuçlarına göre önerdikleri kümeleme modelinin müşteri segmentasyonu için iyi sonuçlar verdiğine değinmişlerdir.

Namyar ve diğerleri [22], yaptıkları çalışmada MİY'in ilk adımı olan "Müşteriyi Tanıma" adımı için segmentasyonun öneminden bahsetmiş olup veri madenciliği araçlarından olan K-Ortalamalar metoduyla müşteri segmentasyonu için yenilikçi bir yaklaşım geliştirmeyi amaçlamışlardır. Girdi değişkenleri olarak demografik değişkenler, RFM ve LFV (lifetime value) kullanılmıştır. Bu yeni yaklaşım iki aşamadan oluşmuş olup ilk adımda K-Ortalamalar kümelemesi ile müşteriler RFM değerlerine göre farklı segmentlere ayrılmışlardır. İkinci adımda ise demografik verilere göre her küme yeniden yeni kümelerle bölünmüştür. Geliştirilen yöntem bir bankanın verileri üzerinde kullanılmış olup mevcut müşteriler ortak işlem davranışlarına ve özelliklerine göre dokuz müşteri grubuna ayrılmıştır. Bu iki adımdan sonra ise LTV kullanılarak müşteriler için bir profil oluşturulmuştur. Sonuç bölümünde bu profiller incelenerek her küme için stratejiler oluşturulabileceği eklenmiştir.

Chan ve diğeri [23], yaptıkları çalışmada çevrimiçi satış kanallarının karlılıklarını artırmak için fiyatlandırma ve promosyon stratejileri oluşturmayı amaçlayarak müşterileri segmente etmişlerdir. Müşterileri segmente edebilmek için RFM verileri kullanılmıştır. Daha sonra segmentasyon sonucunda elden edilen gruplara uygun ürünlerin liste fiyatları ayarlanmış ve müşterilere sunulmuştur. Bu kısımda müşterilerin fiyat üzerinde pazarlık yapmaları ve müşteri bazında hesaplanan indirimli fiyatları sunmak da hedeflenmiştir. Yani, mağaza için daha değerli olan müşterilere daha düşük bir liste fiyatı sunulmuş ve daha sonraki pazarlık sürecinde bu müşterilere daha fazla fiyat tavizi verilmesi amaçlanmıştır. Son olarak da müşterilere satın alma eğilimlerini pekiştirmek için promosyon seçenekleri sunulması önerilmiştir. Önerilen model bir perakende çevrimiçi mağazasında uygulanmıştır ve yaklaşımın performansını değerlendirmek için uygulama öncesi ve sonrası işlem verileri karşılaştırılmıştır.

Lee [24], yaptığı çalışmada bir hastanenin sadık hastalarını bulabilmek ve tıbbi hizmet kullanım örüntülerini ortaya çıkarabilmek için RFM verileriyle segmentasyon çalışması yapmıştır. RFM değişkenlerinin değerlerine göre müşterileri segmente ederek iki gruba ayırmıştır ve önemli ölçüde tıbbi hizmet kullanan/harcama yapan müşterileri sadık müşteriler olarak adlandırmıştır ve bunları hedef pazar olarak belirlemiştir. Daha sonra Karar Ağacı metodu kullanılarak sadık müşterilerin tıbbi hizmet kullanım örüntülerini modellemiştir. Karar ağacı sonucunda sadık müşterilerin tahmin edilebilir faktörleri; kalış süresi, seçilebilir tedavinin kesinliği, ameliyat, eşlik eden tedavi sayısı, hasta odası türü ve taburcu edildikleri bölüm olarak belirlenmiştir. Çalışmanın sonuç bölümünde ise MİY'de SHK stratejisi (segmentasyon, hedefleme ve konumlandırma) ve RFM modelini veri madenciliği ile birleştirmenin pratik kullanımından bahsedilmiştir.

Ezenkwu ve diğeri [25], müşterilerin her birinin ihtiyaçlarının anlaşılmasının ve müşterilerin her birine özel hizmet seviyelerinin ve pazarlama stratejilerinin geliştirilmesinin gerekliliğinden bahsederek bu anlayışın ancak sistematik müşteri segmentasyonu ile elde edilebileceğini söylemişlerdir. Çalışmada Nijerya'da birçok şubesi olan bir mega işletme biriminden alınan veriler ile müşteri segmentasyonu için K-Ortalamalar kümeleme algoritmasının bir MATLAB uygulamasını sunmuşlardır. Kümelemede müşterinin aylık ortalama satın aldığı mal miktarı ve aylık ortalama müşteri

ziyareti sayısı olarak iki özellik dikkate alınmıştır. Birkaç yinelemeden sonra %95 doğrulukla 4 küme yani müşteri segmenti belirlemiştir. Bu kümeleri de Yüksek Alıcı-Düzenli Ziyaretçiler, Yüksek Alıcı-Düzensiz Ziyaretçiler, Düşük Alıcı-Düzenli Ziyaretçiler ve Düşük Alıcı-Düzensiz Ziyaretçiler olarak adlandırmışlardır. Ayrıca sonradan gelecek bir verinin en yakın benzerliğe sahip olduğu herhangi bir kümeye atanabileceğini eklemiştir.

Başkol [26], yapmış olduğu çalışmada yerli bir perakende firmasının verilerini kullanarak, RFM modeli yardımıyla müşteri segmentasyonunun iyi bir şekilde yapılabileceğini göstermeyi ve ortaya çıkan kümelerin satın alma davranışları ile nasıl bir ilişki içerisinde olduklarını ortaya çıkarmayı amaçlamıştır. Bu perakende firmasının sistemine kayıtlı (sadalet kartı kullanarak kaydolmuş olanlar) 18.975 müşterinin alışveriş kayıtları veri olarak kullanılmıştır. İlk önce bu müşteriler RFM modeli ile segmente edilmiş olup elde edilen kümelere isimler verilmiştir. Daha sonra segmente edilen bu müşterilerin hangi markalara daha yakın olduklarını belirleyip marka seçimlerini görselleştirerek perakendecilere daha ayrıntılı bir biçimde müşteri segmentlerini inceleyebileceklerini göstermeye çalışmıştır. Çalışmada müşteri gruplarının sıklıkla tercih ettikleri makarna markaları tespit edilmiş olup grupların hangi markalarla ilişkilendirildiğini bulabilmek için uyum analizi uygulanmıştır.

Wei ve diğerleri [27], yaptıkları derleme makalesinde RFM modeliyle alakalı birçok soruya cevap verebilmeyi amaçlayarak modelin uygulanması hakkında kapsamlı bir inceleme sunmuşlardır. Özellikle “RFM modelinin tanımları ve puanlama şemaları nelerdir?”, “RFM modeli nasıl uygulanır?”, “RFM modelinin avantajları ve dezavantajları nelerdir?”, “RFM modelinin ve diğer modellerin görelî avantajları ve dezavantajları nelerdir?”, “RFM diğer değişkenler veya diğer modellerle nasıl birleştirilir?” sorularına cevap vermişlerdir. Sonuç bölümünde modelin çok çeşitli müşteri ihtiyaçlarını karşılamak için etkili stratejiler geliştirebileceğini ve kâr amacı gütmeyen kuruluşlar ve devlet kurumları için dahi etkili pazarlama stratejisi geliştirme konusunda faydalı olabileceğini eklemiştir.

Rachid ve diğçerlerinin [28] yapmış oldukları çalıřmanın amacı, LRFM modeli ve kümeleme analizi tekniklerini birleřtirerek, e-ticaret alanında müşterilerin segmentasyonu kapsamında bir metodoloji önerebilmektir. Fas'ta yer alan bir çevrimiçi satış řirketinin müşterilerinin sadakatini belirleyebilmek için LRFM yönteminin sonuçlarını kullanmışlardır. Veri olarak belirli bir aralıkta bu řirketin web sitesinden alışveriş yapmış olan 730 müşterisinin verileri kullanılmıştır. Çalıřma iki aşamadan oluşmaktadır. İlk aşamada kendi kendini düzenleyen haritalar metodu (SOM) yardımıyla en iyi küme sayısını bulabilmek ve başlangıç merkezini belirlemek amaçlanmıştır. Bu kısımda verinin 9 kümeye ayrılması gerektiđi belirlenmiştir. İkinci aşamada bu 730 kişilik müşteri verisi LRFM deđerlerine bakılarak K-Ortalamlar algoritmasıyla 9 kümeye ayrıştırılmıştır. Sonuç bölümünde, küme 6'nın L, R, F ve M ortalama deđerlerinin, genel ortalama deđerden yüksek olmasından dolayı bu küme en önemli küme olarak belirlenmiştir. Ayrıca tüm kümelerin analizi yapıldıktan sonra, çevrimiçi alışveriş sitesiyle müşterileri arasındaki iliřkinin geliştirilebilmesi amaçlanarak bazı pazarlama stratejileri önerilmiştir.

Yousefi ve diğçerleri [29], yaptıkları derleme çalıřmasında literatürde bulunan hiyerarşik kümeleme, bölümsel kümeleme, yoğunluk tabanlı kümeleme, ızgara tabanlı kümeleme ve model tabanlı kümeleme yöntemlerinin birbirinden farklılıkları, avantajları ve dezavantajları tartışılmıştır. Algoritmaların tarihçelerinden, kullanım alanlarından bahsedilmiş ve benzerlik mesafe ölçüm yöntemleri incelenmiştir.

3. VERİ MADENCİLİĞİ VE KÜMELEME ANALİZİ

Bu bölümde veri madenciliğinin tanımı, amacı, uygulama alanları, tarihçesi, yöntemleri, bilgi keşfi süreci, makine öğrenmesi tanımı, makine öğrenmesi yöntemleri ve son olarak tez çalışması kapsamında kullanılan yöntemlerden bahsedilmiştir.

3.1. Veri Madenciliği

Bu bölümde veri madenciliğinin tanımı, amacı, uygulama alanları, tarihçesi, yöntemleri ve bilgi keşfi sürecinden bahsedilmiştir.

3.1.1. Veri Madenciliği Tanımı ve Amacı

Veri, niteliksel veya niceliksel değişkenlerden oluşan, henüz işlenmemiş (ham) veya işlenmiş bir veya birden fazla bilgiden oluşan kümedir. Genellikle gözlem, deney, araştırma, sayım veya ölçüm yollarıyla elde edilir ve analiz edilmediği takdirde genel olarak bir anlam ifade etmez.

Veri madenciliği hakkında birbirine yakın olan birçok tanım yapılmıştır. Meyer ve Canon [30], gelecekle ilgili tahminler yaparak belirsizlikleri aydınlatılabilmek amacıyla karışık ve çok büyük miktardaki verilerin içerisinde işe yarayan ve birbirleri ile anlamlı olan bilgilerin öne çıkarılıp, bu bilgiler üzerinde belirli araçlar kullanılarak yapılan analizlerin sonucunda yararlı bilgilere ulaşılması şeklinde bir veri madenciliği tanımlaması yapmıştır. Ganesh [31], veri madenciliğini çok sayıda ve karmaşık halde bulunan büyük bir veri kümesinin içerisinde gizli olan öz bilginin ortaya çıkarılması süreci olarak tanımlamıştır. Rahman ve arkadaşları [32], veri madenciliğini büyük veri kümesinden değerli bilgi ve/veya kalıpları çıkarmak olarak tanımlamaktadır.

Veri madenciliği, bilgisayar programları aracılığıyla büyük veri ve çoklu değişkenlerle ilgilenerken anlamlı ve değerli bilgilerin ortaya çıkarılması ve analiz edilmesi olarak da tanımlanabilir. Veri madenciliğini bir süreç olarak ele almak gerekirse, büyük veri

kümelerinin analize uygun hale getirilmesi, uygun metotların seçilmesi, uygulanması ve sonuçların anlamlandırılması olduğu söylenebilir. Veri madenciliğinin amacı genel olarak veriyi sınıflandırmak, kümelendirmek, tahmin yapabilmek ve benzer grupları tanımlayabilmektir. Teknolojinin geldiği nokta düşünüldüğünde verinin elde edilmesi, kayıt altına alınması ve veriye erişilmesi kolay bir durumdur. Fakat veriyi anlamak ve anlamlandırmak aşamasına gelindiğinde zorluk başlamaktadır. Veri madenciliği bu noktada devreye girmektedir.

Veri madenciliği istatistik, yapay zekâ ve makine öğrenmesi olarak 3 farklı disiplinden oluşmaktadır. İstatistik veriler arasındaki sayısal ilişkilerin ortaya çıkarılmasını, yapay zekâ ilgili yazılımlar aracılığıyla insan benzeri karar verme yapılarının ortaya çıkarılmasını, makine öğrenmesi ise büyük veriler üzerinde çalışarak değerli bilgileri edinmeyi amaçlamaktadır [33].

3.1.2. Veri Madenciliği Uygulama Alanları

Veri madenciliği, bilgi keşfi sürecinde farklı disiplinlerin bir araya getirildiği multidisipliner bir yaklaşımdır. İstatistik, algoritmalar, veri tabanı sistemleri, makine öğrenmesi, veri görselleştirme ve diğer disiplinlerdeki teknikleri bir araya getirerek verilerden yararlı bilgiler edinmemizi sağlamaktadır. Bu yüzden hemen hemen her sektörde kullanılmaktadır. Özellikle pazarlama stratejilerinin belirlenmesinde oldukça sık kullanılmaktadır [34].

Büyük veri kullanımı farklı sektörlerde giderek artmakta ve araştırmacılar da bu kavrama yoğun ilgi göstermektedir. Özellikle 21. yüzyılda veri ambarlamasında firmalar tüketicilerden, tedarikçilerinden, ortaklarından ve rakiplerinden oldukça ayrıntılı veriler toplamaktadır. Bu verileri de birçok alanda kullanabilmektedirler. Veri madenciliğinin kullanım alanları yüzlerce olmakla birlikte en çok kullanılanları şu şekilde sıralanabilir:

- Pazarlama ve Müşteri İlişkileri Yönetimi
 - Müşteri özellikleri ve davranışlarının belirlenmesi

- Satış tahminleri
 - Pazarlama kampanyalarını doğru kanala yönlendirerek karlılığın maksimize edilmesi
 - Müşteri sadakatinin artırılması üzerine çalışmalar
- Bankacılık ve Finans Sektörü
 - Harcama verilerine göre müşteri segmentlerinin belirlenmesi
 - Kredi taleplerinin değerlendirilmesi
- Elektronik Ticaret
 - Web sayfalarına veya mobil uygulamalara yapılan müşteri ziyaretlerin takibi ve analizi
 - Kullanıcı davranışlarına göre web arayüzlerinin geliştirilmesi
- Mühendislik, Sigortacılık, Sosyal Medya, Sağlık ve İlaç Sektörü, Ticaret, Spor Bilimleri, Gayrimenkul, Eğitim, Telekomünikasyon, Satış, Finans, İmalat, İnsan Kaynakları Yönetimi, Güvenlik ve İstihbarat vb.

3.1.3. Veri Madenciliği Tarihçesi

Çok hızlı bir şekilde gelişim gösteren teknoloji sayesinde veri madenciliğinin önemi oldukça iyi bir şekilde anlaşılmış olup, her geçen saniye eklenen veriler sayesinde günümüzde sonu olmayan devasa bir veri havuzu oluşmuştur.

Veri madenciliği kavramını gelişen teknoloji sayesinde günlük hayatta dahi birçok kez duyduğumuz için yeni bir kavram olarak düşünülebilir. Ancak veri madenciliğinin tarihi geçmişe dayanır. Geçmişten beri insanlar geleceğin belirsizliğinin yaratabileceği olumsuzlukları önleyebilmek, tedbir alabilmek amacıyla gelecek ile ilgili tahminlerde

bulunmak ve görüş alanlarını aydınlatmak istemişlerdir. Bundan dolayı geçmişte de şimdiye göre daha az modern olsa bile veri madenciliğinin kullanıldığından söz edilebilir.

1700'lü yıllarda Bayes Teoremi ve 1800'lü yıllardaki regresyon analizi ile erken veri madenciliği metodu başlamıştır. Büyük verinin, her ne kadar son yıllarda dijitalleşmenin etkisiyle ortaya çıktığı düşünülse de kökeni, veri merkezlerinin kurulduğu 1960'lı yıllara kadar dayanmaktadır ve günümüzde farklı birçok alanda kullanımı ve yayılımı artmaktadır [35].

1960'lı yıllarda veriler, bilgisayar ortamında saklanmaya başlanmış bu veriler üzerinden analizler yapılmaya başlanmıştır. O dönem için bilgisayardaki veriler üzerinde yeterince uzun bir süre çalışma yapıldığında istenilen sonuçlara ulaşabileceği gözlemlenmiştir. O zamanlar bu işleme veri madenciliği yerine “veri tarama”, “veri yakalama” gibi isimler verilmiştir.

Bilim insanlarının “İlişkisel Veri Tabanı Yönetim Sistemleri” uygulamalarını kullanmaya başlamaları ile birlikte 1970'li yıllarda basit kurallara dayanan sistemler kurmuşlar ve basit biçimde makine öğrenimini sağlamışlardır. Bu gelişme ile birlikte 1980 yıllarında bu sistemler daha çok kullanılmaya başlanmış ve şirketler de bu yararlı keşfi müşterileri, ürünleri ve benzeri gibi fayda sağlayacak konularda verileri içeren veri tabanları oluşturmuşlardır [36].

Yine 1980'li yıllarda analizlerin daha anlık ve hızlı yapılabilmesi için SQL (Select Query Language) veri tabanı sorgulama dili kullanılmaya başlanmıştır. 1989 yılında KDD (IJCAI)-89 Veri Tabanlarında Bilgi Keşfi Çalışma Grubu toplantısı ve ardından sonuç bildirgesi sayılan “Knowledge Discovery in Real Databases: A Report on the IJCAI-89 Workshop” makalesi yayımlanmış, bu makalede yer alan KDD ile ilgili temel kavramlar ve tanımlar veri madenciliğinin gelişim sürecine hız katmıştır [36]. 1990 yılı ve sonrasında Veri Ambarlama ve Karar Destek Sistemleri ortaya çıkmıştır.

2000'li yıllara gelindiğinde ise veri ambarının da kullanımıyla birlikte veri madenciliği yaygınlaşmaya başlamıştır. Her geçen yıl katlanarak artan veri miktarları ve gelişen

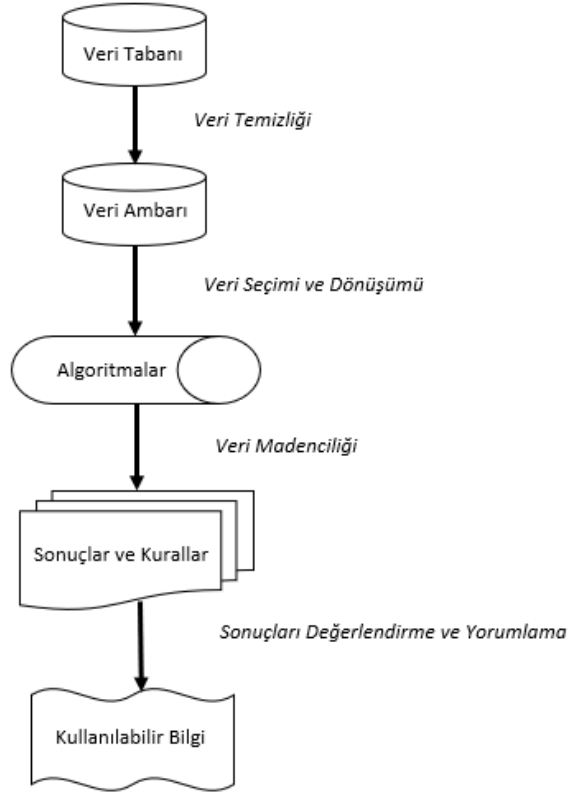
teknolojik yöntemler sayesinde 2000 yılı ve sonrasında gerek teknolojik gelişmeler gerekse milyonlarca insanı etkisi altına alan sosyal mecralar da göz önüne alındığında veri madenciliği büyük bir hızla gelişerek günümüzde bulunduğu noktaya gelmiştir.

3.1.4. Bilgi Keşfi Süreci

Bilgi keşfi süreci için büyük verilerdeki değerli ve anlamlı kalıpları çıkarma sürecidir denilebilir [37]. Büyük veriler üzerinden analiz yapılabilmesi için başlangıçta problemin tanımlanması gerekmektedir. Daha sonra veri madenciliği yani bilgi keşfi süreci; veri toplama, veri hazırlama, model seçimi ve model oluşturma, son olarak değerlendirme ve bilgi aktarımı şeklinde gerçekleşmektedir [38]. Daha açık şekliyle büyük veriden bilgi edinme süreci adımları aşağıdaki gibi sıralanabilir:

1. Verinin temizlenmesi/hazırlanması: Veri grubunda yer alan hatalı, eksik veya tutarlı olmayan verilerin kaldırılması ya da tutarlı hale getirilmesidir.
2. Verinin entegrasyonu: Farklı veri kaynaklarından alınan verilerin bir araya getirilmesidir.
3. Verinin seçimi: Veri grubundan yapılacak çalışmada kullanılacak anlamlı verilerin seçilmesidir.
4. Verinin dönüştürülmesi: Yapılacak çalışma için eldeki verilerin veri madenciliği tekniklerine göre yararlı formatlara dönüştürülmesidir.
5. Veri madenciliği: Veriye akıllı yöntemler uygulanarak veriden bilgi ve örüntü elde edilmesidir.
6. Örüntü değerlendirmesi: Veri madenciliği ile ulaşılan değerli bilgi ve kalıpların ilgili analiz yöntemleriyle değerlendirilmesi ve hedeflenen hipotezlerin geçerli olup olmadığının test edilmesidir.

7. Bilginin sunulması: Ulaşılan bilgilerin grafiklerle görselleştirilmesi ve sonuçları kullanacak ilgililere sunulması işlemleridir [39]. Bilgi keşfi sürecini Şekil 3.1 ile de anlatabiliriz.



Şekil 3.1. Bilgi Keşfi Süreci [40].

3.1.5. Veri Madenciliği Yöntemleri

Veri madenciliği çalışmalarında kullanım için geliştirilen birçok yöntem vardır. Bu yöntemlerin birçoğu istatistiksel tabanlı yöntemlerdir. Özellikle işletmeler son zamanlarda müşteri verilerinden yola çıkarak pazarlama kaynaklarını doğru kanallara yönlendirerek karlılığı ve kaynak verimini maksimuma çıkarmayı hedeflemektedir. Bu tez çalışmasında da veri madenciliği yöntemlerinden faydalanılarak müşteri segmentleri belirlenmeye çalışılacaktır.

Veri madenciliği yöntemlerinin kullanımını elimizdeki verinin türü ve çalışmalar sonunda elde edilecek sonucun kullanım amacına göre farklılaştırır. Bu yöntemleri tahmine dayalı modeller ve tanımlayıcı modeller olarak ikiye ayırabiliriz.

Tanımlayıcı modeller, veri setindeki karar vermeye rehberlik edebilecek ilişkileri ortaya çıkarırlar ve tanımlarlar. Tanımlayıcı modellerde kullanılan teknikler temelde kümeleme ve birliktelik kurallarıdır. Eş zamanlı gerçekleşen ilişkilerin tanımlanmasında birliktelik kuralları kullanılır. Kümeleme ise veri setindeki etiketlenmemiş verileri özelliklerine göre kümelere ayırma işlemidir. Aynı küme içindeki elemanların birbirine benzerliğinin yüksek, diğer kümelerle benzerliğinin düşük olması amaçlanır.

Tahmine dayalı modeller, sonuçları önceden bilinen verilerden bir model geliştirmek ve bu modeli kullanarak bilinmeyen veriler için sonuçları tahmin etmek için kullanılır. Tahmine dayalı modellerde kullanılan teknikler temelde sınıflandırma, regresyon ve zaman serileri analizleridir. Bir veri setindeki etiketlenmemiş verilerin hedef kategorilere veya kümelere atanması sınıflandırma yöntemidir. Bir bağımlı değişken ile bağımsız değişkenler arasındaki ilişkiyi tahminlemeye çalışmak ise regresyon yöntemi ile yapılır.

3.2. Makine Öğrenmesi

Öğrenme kelimesi anlamsal olarak; belli durumlar ve sorunlar karşısında bilgiyi algılama ve bunlara göre tepki ve davranış oluşturma, gerektiğinde bunları değiştirip yenilerini edinebilme yeteneği şeklinde tanımlanmaktadır. Yapılan bu tanımlamaya göre makine öğrenmesi kavramı için karşılaşılan durumlar karşısında makinelerin kendilerini eğiterek kararlar verebilmesini sağlayan algoritmaların geliştirilmesidir.

İlk olarak Amerikan bilgisayar bilimcisi Arthur Samuel tarafından 1959'da bir araya getirilen "makine öğrenmesi" terimi, "bilgisayarın açıkça programlanmadan öğrenme yeteneği" olarak tanımlanmaktadır [41]. Arthur Samuel bu çalışmasında bir bilgisayarın programı yazan kişiden daha iyi dama oyunu oynamasını doğrulamayı amaçlamıştır.

En temelde, girdi verilerini alarak analiz eden ve çıktı değerlerini kabul edilebilir bir aralıkta tahmin etmeye çalışan programlanmış algoritmalarıdır. İnsanlar tarafından rahatlıkla anlaşılabilir, basit sınıflandırıcı ifadeler üretmeyi amaçlar. Makine öğrenmesi genellikle yapay zekâ ile karıştırılan yanlış bir terimdir fakat makine öğrenmesi aslında bir yapay zekâ alanı/alt tipidir. Kısacası aralarındaki ilişki, birbirinden ayrılmaz ikili olmaları ve yapay zekâ genel bir kavramken makine öğrenmesi onun bir alt dalıdır şeklinde açıklanabilir.

Makine öğrenmesine dayalı yöntemler, veri birimleri arasındaki ilişkiyi dikkate alırken aynı zamanda istatistiksel çıkarımlarda bulunarak tespit ve öngörülere dayanarak kararlar alınmasına yardımcı olurlar. Bu yöntemler temelde denetimli ve denetimsiz olarak ikiye ayrılırlar [42]. Denetimsiz öğrenme herhangi bir ön bilgi olmayan verilerin nasıl düzenlendiğini ve organize olduğunu belirlemeye çalışır. Denetimli öğrenme ise önceden tanımlanmış yapılmış kategorilere veya kümelere yeni gelen ikiden fazla kümeyi eşleştiren modeller geliştirmeye çalışır. Temel amaç keşfedilmemiş yapıların tespit edilmesidir [43]. Bu iki öğrenme yöntemine ek olarak yarı denetimli ve takviyeli makine öğrenmesi yöntemleri de vardır.

Makine öğrenmesi, insan tarafından uzun sürede hesaplanabilecek veya yapılabilecek bir problemin veya durumun, bilgisayarlar tarafından çok kısa sürede yapılabilmesini sağlamaktadır. Birçok gerçek dünya problemine makine öğrenmesi teknikleriyle çözüm aranmaktadır. Bunlara örnek verecek olursak:

- Görüntü işleme ve bilgisayarla görme: Yüz tanıma, hareket algılama, nesne algılama.
- Finans: Kredi skorlaması, algoritmik ticaret.
- Biyoloji: İlaç keşfi, DNA dizilimi, tümör bulma.
- Enerji Üretimi: Fiyat ve yük tahmini.
- Otomotiv, havacılık ve üretim: Öngörücü bakım

Makine öğrenmesinin adımları aşağıdaki gibi listelenebilir:

- 1- Verinin yüklenmesi
- 2- Veri ön işleme
- 3- Özellik çıkarımı
- 4- Modeli kurmak ve eğitmek
- 5- Modeli geliştirmek

3.2.1. Makine Öğrenmesi Metotları

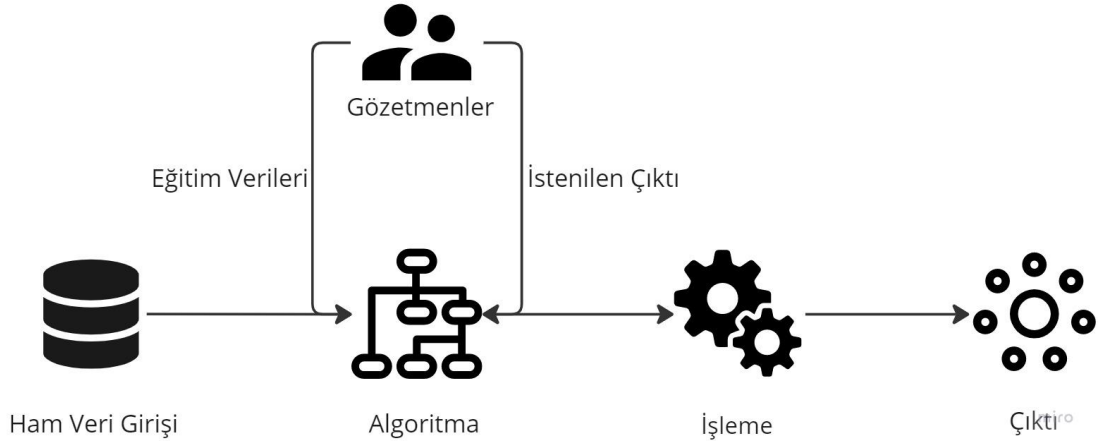
Makine öğrenmesi metotlarının sayısı mevcutta çok fazladır ve her geçen gün de bu sayı artmaya devam etmektedir. Bu çalışmada makine öğrenmesi yöntemlerinden en temelleri anlatılacaktır.

Makine öğrenmesi metotları temelde 4 başlık altında incelenebilir. Bunlar:

- Denetimli Makine Öğrenmesi
- Yarı Denetimli Makine Öğrenmesi
- Takviyeli Makine Öğrenmesi
- Denetimsiz Makine Öğrenmesi

3.2.1.1. Denetimli Makine Öğrenmesi

Denetimli bir öğrenme algoritması bilinen bir girdi verisi seti ve verilere bilinen yanıtları almaktadır ve ardından yeni verilere yanıt için makul tahminler oluşturmak üzere bir modeli eğitmektedir. Girdi verileri modeli besledikçe, modelin uygun şekilde yerleştirilmesini sağlayan bir takviye öğrenme prosesiyle ağırlıkları ayarlamaktadır [44]. Yani Denetimli Öğrenme için bir çeşit organize sistemi denilebilmektedir ve algoritmaların temel adımları Şekil 3.2'deki gibi gösterilebilmektedir.



Şekil 3.2. Denetimli Makine Öğrenmesi Süreci [45].

Genellikle sınıflandırma, regresyon ve tahmin problemlerinde kullanılırlar.

Sınıflandırma: Makine öğrenmesi ve istatistikte; sınıflandırma, bilgisayar programının verilen veri girişinden öğrendiği ve sonrasında yeni gözlemleri sınıflandırmak için bu öğrenmeyi kullandığı denetimli öğrenme yaklaşımıdır. Genel olarak ikili veya çoklu sınıflandırma şeklinde ikiye ayrılmaktadır. İkili sınıflar 0/1, doğru/yanlış, ret/kabul gibi tanımlanırken, çoklu sınıflar ise ikiden fazla durumları içeren sınıfları tanımlar.

Sınıflandırma yöntemlerinden bazıları aşağıdaki gibi sıralanabilir.

- Naive Bayes
- K-En Yakın Komşu
- Lojistik Regresyon
- Karar Ağacı/Şeması
- Rastgele Orman
- Destek Vektör Makineleri
- Diskriminant Analizi
- Yapay Sinir Ağları

Regresyon: Regresyon analizinde, makine öğrenme programı değişkenler arasındaki ilişkileri tahmin etmeli ve anlamalıdır. Genellikle neden-sonuç ilişkileri araştırılır. Yani

değişkenlerden biri ya da birkaçının, diğer bir ya da birkaç değişkeni ne ölçüde etkilediği incelenir. Bu, özellikle tahmin için faydalıdır. Değişkenler arasında ilişki olduğu gözlemlenirse ilişkinin derecesi matematiksel bir fonksiyon olarak ortaya konur. Bu fonksiyona regresyon fonksiyonu denir.

Regresyon algoritmalarından bazıları şu şekilde sıralanabilir:

- Basit Doğrusal Regresyon
- Çoklu Doğrusal Regresyon
- Polinom Regresyon
- Lojistik Regresyon
- Destek Vektör Regresyonu
- Karar Ağacı Regresyonu
- Rastgele Orman Regresyonu

Tahmin: Tahmin, geçmiş ve şimdiki verilere dayanarak gelecekle ilgili tahminler yapma sürecidir ve genellikle eğilimleri analiz etmek için kullanılır.

3.2.1.2. Yarı Denetimli Makine Öğrenmesi (Semi-supervised)

Denetimli ve denetimsiz makine öğrenmesine benzer özellikleri bulunan bir algoritmadır. Benzeme sebeplerinden en önemlisi denetimli öğrenmede kullanılan etiketli verileri, denetimsiz öğrenmede ise etiketsiz verileri içerisinde barındırmasıdır. Yarı denetimli makine öğrenmesi algoritmalarından bazıları şu şekilde sıralanabilir:

- Kendi Kendine Eğitim
- Üretken Karışım Modelleri
- Yarı Denetimli Destek Vektör Makinesi
- Grafik Tabanlı Algoritmalar
- Çoklu Görüntü Algoritmaları

3.2.1.3. Takviyeli Makine Öğrenmesi

Takviyeli makine öğrenmesi, bulunduğu ortamı algılayan ve o ortamda hareket eden özerk bir ajanın, hedefine ulaşabilmesinde doğru kararlar almayı nasıl öğrenebileceğini gösterir [46]. Kesin bilgilerin yeterli olmadığı durumlarda bile hiçbir ön bilgiye ihtiyaç duymadan, yani elimizde bir veri olmadan da süreci başlatabilen bir algoritmadır. Bu algoritma kesinlikten ziyade deneme-yanılma yöntemiyle çalışır. Takviyeli makine öğrenmesinin denetimli makine öğrenmesinden en belirgin farkı burada zaman önemli bir faktör olmasıdır. Robotik, oyun programlama, hastalık teşhisi ve tanı koyma, otomasyon gibi sistemlerde yaygın olarak kullanılmaktadır. Takviyeli makine öğrenmesi algoritmalarından bazıları şu şekilde sıralanabilir:

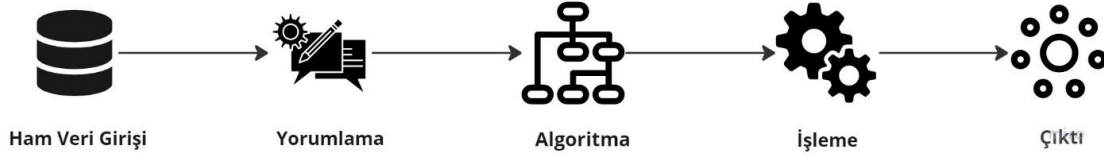
- Zamansal Fark Öğrenimi Algoritması
- Q Öğrenme Algoritması
- Sarsa Algoritması

3.2.1.4. Denetimsiz Makine Öğrenmesi

Denetimsiz öğrenmede, daha önce etiketlenmemiş verilerin algoritmayla kümelendirilmesi ve hangi verinin hangi sınıfa ait olduğunun belirlenmesi amaçlanır [47]. Denetimli makine öğrenmesinin aksine, örnekler etiketsiz olarak sağlanır ve elemanlar arasındaki ilişkiyi veya veri yapısını incelemektir. Denetimsiz makine öğrenmesinde, denetimli makine öğrenmesinde olduğu gibi y bağımlı değişkeni bulunmamaktadır, yani veri etiketsizdir [48]. Yani burada bağımsız değişkenler yer almaktadır ve bunlar herhangi bir bağımlı değişken ile ilişkilendirilemediği için veri setlerinde öğrenme işlemleri yapılmaktadır.

Denetimsiz öğrenme algoritmaları veri setinde analizler yaparak gözle tespit edilemeyen ilişki ve örüntüleri ortaya çıkartmaktadır [49]. Denetimsiz algoritmaların temel adımları Şekil 3.3'deki gibi gösterilebilmektedir.

- Bilinmeyen Çıktı
- Eğitim Verileri Yok



Şekil 3.3. Denetimsiz Makine Öğrenmesi Süreci [45]

Denetimsiz makine öğrenmesinin iki önemli konusu vardır. Bunlar birliktelik kuralları (ilişkilendirme) ve kümelemedir.

Birliktelik Kuralları: Birliktelik kuralı alışverişlerde sıklıkla karşımıza çıkmaktadır. Çevrimiçi alışveriş siteleri alışveriş verilerini birliktelik kuralına göre analiz ederler. Bir veya birden fazla alışveriş yapan müşterilerin tercih ettikleri ve satın almaya meyilli oldukları ürün veya hizmetlerin belirlenmesi için birliktelik kuralları üzerinden analizler yapılır. Bu analizler sonucunda müşterilere ilişkili ürünler önerilerek daha fazla ürün satın alınması hedeflenmektedir [50]. Bu öneriler, müşteri sepetine bir ürün eklediğinde aldığı ürünle çoğunlukla ilişkili başka bir ürünü “bu ürünü de sepetinize eklemek ister misiniz” sorusu sorularak gerçekleşir. Bu önerinin yapılabilmesi öncesinde yapılan analizlerle gerçekleşebilir.

Birliktelik kurallarında kullanılan farklı algoritmalar bulunmaktadır ve bunlar aşağıda sıralanmıştır. Aralarında kullanımı en sık tercih edilen Apriori algoritmasıdır.

- AIS
- Apriori
- CHARM
- FP-Growth
- Partition
- RARM
- SETM

Kümeleme Algoritmaları: Kümeleme, aynı gruptaki veri noktalarının diğer veri noktalarına daha benzer ve farklı gruptaki veri noktalarından farklı olacak şekilde birkaç gruba ayırma algoritmasıdır. Temel olarak, etiketlenmemiş verileri, aralarındaki benzerlik ve farklılık temelinde analiz ederek gruplara ayırmayı amaçlar. Sınıflandırma yönteminden ayrılan özelliği öncesinde bilinen yani etiketlenmiş grupların olmamasıdır. Kümeleme algoritmaları istatistiksel veri analizi ve örüntü tespiti gibi alanlarda sıklıkla kullanılmaktadır.

Kümelemenin mantığında verilerin var oldukları küme içerisindeki benzerliklerini en üst düzeye çıkarma, diğer kümelerle arasındaki benzerliği de en aza indirme durumu vardır. Yani yer aldıkları küme içerisinde homojen, diğer kümelerle heterojen özellik göstermeleri beklenmektedir [51]. Literatürde de birçok tanımlaması bulunmaktadır. Eldeki veri setinin hangi öznitelikler kullanılarak kaç farklı gruba ayrılması gerektiği, hangi verinin hangi gruba ayrılacağı gibi konular eldeki veriye göre belirlenmekte ve belirlenen her bir veri grubu küme olarak adlandırılmaktadır [52].

Kümelere ayrılacak veriler, vaka, varlık, nesne veya örüntü olarak ifade edilmektedir. Bu verilerin benzerliklerini tespit etmek amacıyla kullanılan özellikleri ise değişken, öznitelik veya karakter olarak adlandırılmaktadır [53].

Özellikle gelişen teknoloji ve günümüzde veri sayılarının artmasıyla birlikte, verilerden faydalı bilgilerin elde edilmesi amacıyla kümeleme analizleri sıklıkla kullanılmaktadır. Özellikle tıp, biyoloji, telekomünikasyon, bankacılık, ekonomi, psikoloji gibi alanlarda kullanım oranları fazlasıyla yüksektir [54]. Yaygın olarak kullanım alanlarından biri olan ve bu tez çalışmasının da içeriğini oluşturan müşteri verileriyle pazar bölümlendirme yapılması ve elde edilen gruplara uygun pazarlama stratejilerinin geliştirilmesi de özellikle son yıllarda fazlasıyla tercih edilen çalışma alanlarından biridir [55]. Kümeleme analizi, ortak amacı verilerin benzerlerinden oluşan kümelerini tespit etmek olan numerik yöntemleri temsil eden genel bir terimdir [56]. Kümeleme algoritmalarında kullanılan yöntemleri aşağıdaki gibi sıralayabiliriz:

- Merkez Tabanlı Kümeleme Yöntemleri
 - K-Ortalamalar Kümeleme
 - K-Medoids Kümeleme
 - K-Modes Kümeleme
 - PAM
 - CLARA
 - CLARANS

- Hiyerarşik Kümeleme Yöntemleri
 - BIRCH
 - CHAMELEON
 - ROCK
 - CURE
 - SLINK

- Yoğunluk Tabanlı Kümeleme Yöntemleri
 - DBSCAN
 - OPTICS
 - DENCLUE
 - DBCLASD

- Dağıtım Tabanlı Kümeleme Yöntemleri
 - Gauss Karma Modeli

- Model Tabanlı Kümeleme Yöntemleri
 - Kendini Düzenleyen Haritalar (SOM)
 - COBWEB

- CLASSFIT
- MCLUST
- EM

- Izgara (Grid) Tabanlı Kümeleme Yöntemleri
 - STING
 - WAVE Cluster
 - CLIQUE

- Diğer Kümeleme Yöntemleri
 - Mean Shift

3.3. Tez Çalışması Kapsamında Kullanılan Yöntemler

Bu bölümde, tez çalışması kapsamında yapılacak uygulamanın adımlarında kullanılacak yöntemlerden bahsedilecektir.

3.3.1. Öznitelik Seçimi

Büyük veri içerisinde yer alan boyutlar günden güne artış göstermektedir. Bu artış sebebiyle faydalı bilgiye ulaşmak zorlaşmaktadır ve bu yüzden hesaplama maliyetlerinde artış ve performanslarda kayıplar yaşanmaktadır. Bu sorunların azaltılması veya önlenmesi için öznitelik seçimi önemli bir adım olmaktadır. Çünkü makine öğrenmesi tekniklerinde modeli eğitmek için kullanılan veri özniteliklerinin model performansına etkisi çok yüksektir. Öznitelik seçme adımının amacı verilerdeki boyut sayısını azaltarak faydalı bilgiye kayıp yaşanmadan hızla ulaşılabilmesini sağlamaktır.

Öznitelik seçimi, problemin sonucunu en çok etkileyen özelliklerin otomatik veya manuel olarak seçildiği süreçtir. Yanlış öznitelikler ile çalışılması, kurulan modellerin doğruluk oranının azalmasına ve modelin yanlış özniteliklere göre eğitilmesine sebep olabilir [57].

Budak [58], öznitelik seçiminin avantajlarını şu şekilde sıralamaktadır;

- Veri kümesinin boyutunu azaltır ve çalışma hızını artırır,
- Problem çözümüne katkısı en az olan veriyi çıkartır,
- Verinin daha anlaşılır ve basit hale getirilmesini sağlar,
- Depolama için gereken alan azalır.

Çevrimiçi hızlı tüketim ürünleri siparişi veren kullanıcıların segmentasyonu için hangi özniteliklerden faydalanılacağına yanıt bulmak bu tez çalışmasının önemli bir amacıdır. Bu amaç kapsamında bazı metotlardan faydalanılacaktır. Bunlar Calinski-Harabasz İndeksi ve Davies-Bouldin İndeksi'dir.

3.3.1.2. Calinski-Harabasz İndeksi

Bu yöntem 1974 yılında Calinski ve Harabasz tarafından keşfedilmiştir. Bu indeks aşağıdaki gibi hesaplanmaktadır:

$$ch = \frac{tr(B_k)*(N-k)}{tr(W_k)*(k-1)} \quad (3.1)$$

Bu formülde N veri sayısını, k küme sayısını ifade etmektedir. $tr(B_k)$ küme içi uzaklık karelerinin toplamını, $tr(W_k)$ ise kümeler arası uzaklık karelerinin toplamını ifade etmektedir. Calinski-Harabasz İndeksi'nde en yüksek ch değeri tercih edilir [59]. Calinski-Harabasz İndeksi öznitelik seçimi, kümeleme modellerinin performanslarının karşılaştırılması gibi amaçlarla kullanılmaktadır. Tez çalışmamızda da iki adımda da kullanımı tercih edilmiştir.

3.3.1.3. Davies-Bouldin İndeksi

Kümeler içerisinde yer alan verilerin merkez noktasına olan uzaklığını minimum, elde edilen kümeler arası uzaklığı ise maksimum yapmayı amaçlayan bir yöntemdir. Bu yöntem ile kümeleme performansını ölçmek için aşağıdaki eşitlik kullanılır:

$$DB = \frac{1}{k} \sum_{i=1}^k R_{ij} \quad (3.2)$$

$$R_{ij} = \max_{i \neq j} \frac{s_i + s_j}{d_{ij}} \quad (3.3)$$

“ k ” küme sayısı olmak üzere, $i = 1, 2, \dots, k$ ve $j = 1, 2, \dots, k$ olmak üzere i . ve diğer j kümeleri arasındaki maksimum karşılaştırma oranı her bir küme için R_{ij} ile hesaplanır. d_{ij} , i ve j kümelerinin merkezleri arasındaki uzaklığı ifade etmektedir. s_i ve s_j değerleri ise küme elemanlarının, bulunduğu kümenin merkezlerine olan ortalama mesafesidir. Davies Bouldin İndeksi’nde küçük DB değerleri tercih edilir [60]. Calinski-Harabasz İndeksi gibi Davies-Bouldin İndeksi de öznitelik seçimi, kümeleme modellerinin performanslarının karşılaştırılması gibi amaçlarla kullanılmakta olduğu için de iki adımda da kullanımı tercih edilmiştir.

3.3.2. Küme Sayısının Belirlenmesi

Kümeleme yöntemlerinde eldeki veri setinin sahip olduğu özniteliklere göre kaç adet kümeye ayrılması gerektiği önemli bir sorudur. Özellikle K-Ortalamalar kümeleme yönteminde k sayısı daha çok model kurucusunun önsezisine bırakılır. Bu noktada tez çalışması kapsamında müşteri verileri kullanılan şirketin yetkilileriyle istişare edilecektir. Bu görüşme öncesinde de Elbow metodu ile bir sonuç bulunacak ve tartışmaya sunulacaktır.

3.3.2.1. Elbow Yöntemi

Elbow yöntemi, her bir noktanın oluşan küme merkezlerine olan uzaklığının karelerinin toplamı (Within Clusters Sum of Square-WCSS) ile hesaplanmaktadır. Bu yöntemde göre WCCS'deki değişim oranının gözlemlendiği grafik çizilir ve eğimin azaldığı veya bir başka deyişle değişim miktarının azaldığı nokta dirsek noktası olarak seçilir. Bu dirsek noktası en iyi k küme sayısını ifade etmektedir [61].

3.3.3. Kümeleme Analizi

Bu tez çalışması kapsamında tespit edilen öznelikler ve k küme sayısı ile merkez tabanlı kümeleme yöntemlerinden olan K-Ortalamlar Kümeleme metodu, dağıtım tabanlı kümeleme yöntemlerinden Gauss Karma Kümeleme metodu ve hiyerarşik kümeleme yöntemlerinden Birch Hiyerarşik Kümeleme metodu kullanılacaktır.

3.3.3.1. K-Ortalamlar Kümeleme

MacQueen tarafından [62], N boyutlu bir anakütleli k adet kümeye ayırmayı amaçlayan bir yöntem olarak tanıtılmıştır.

K-ortalamlar kümeleme, veri seti üzerinden belirli sayıda kümeyi (k adet) gruplamak için geliştirilmiş sade ve basit bir algoritmadır. K-ortalamlar algoritması çoğu büyük/küçük veri kümesi için ideal bir algoritmadır. " k " önceden bilinen ve kümeleme işlemi bitene kadar değeri değişmeyen bir sabit olmalıdır [63].

K-ortalamlar kümeleme, istatistiksel olarak benzer özellikteki verileri aynı kümeye yerleştirir. Bir veri yalnızca bir kümeye ait olabilir. Küme merkezi, kümeyi temsil eden değerdir. " k " parametresi analizini yapacağımız veri setinin kaç adet kümeye ayrılacağını belirler. Bu parametre araştırmacının ön bilgisine ve tecrübesine dayanarak veya çeşitli analiz yöntemleriyle seçilir. " k " sayısının değişimiyle farklı sayıda ve nitelikte kümeler ortaya çıkabilir. En sık tercih edilen kümeleme yöntemlerinden biridir.

Algoritmanın iki amacı vardır. Eldeki “ k ” parametresine göre oluşturulan k adet kümenin içerisindeki verilerin birbirlerine olabildiğince benzemesi ve her kümenin birbirinden olabildiğince farklı olmasıdır. Yöntemin adımları:

- Adım 1: Nesneleri al, küme sayısını belirle ve başlangıç kitle merkezlerini (centroid) belirle. (Centroid tipik olarak kümedeki noktaların bir ortalamasıdır.)
- Adım 2: Her nesneyi en uygun gruba ata ve her atama işleminden sonra atama yapılan k kitle merkezini hesapla.
- Adım3: Yeni oluşan grubu önceki grup ile kıyasla. Grupta değişim gözlemlenmedi ise algoritmayı sonlandır, aksi takdirde adım 2’ye geri dön. Yani nesnelerin kümelemesinde değişiklik olmayana kadar bu işlemler tekrarlanır [64].

K-ortalama algoritmasının performansının, çeşitli mesafe ölçütleri ile birçok veri setinde denenmesiyle, kullanılan veri tabanına ve mesafe ölçütlerine bağlı olduğu sonucuna varılmıştır [65].

3.3.3.2. Birch Hiyerarşik Kümeleme

Birch hiyerarşik kümeleme algoritması, Zhang ve arkadaşları tarafından [66], çok miktarda olan sayısal verilerin kümelenebilmesi için tasarlanmış bir algoritmadır. Bu alanda önerilen ilk algoritma olma özelliği taşımaktadır. Çok adımlı kümeleme çalışması yapmak için uzaklık matrisi yerine kümelerin özelliklerini temsil eden bir ağaç oluşturularak bu ağaç üzerinden kümeleme yapılır. Kümeleme özellik ağacı, kümeler hakkındaki bilgileri özetleyen üç boyutlu bir vektöre denilmektedir.

Model, kümeleme için veri setinde yer alan her veri noktasının eşit öneme sahip olmadığı düşüncesini dikkate alarak verilerin yoğun olduğu bölgeyi toplu olarak tek bir küme olacak şekilde değerlendirilirken, daha seyrek alanlardaki verileri aykırı değer olarak görür ve analiz yapan kişinin tercihine göre veri setinden çıkarmasına izin verir. Bunun yanı sıra sadece dairesel kümeleri bulabilir. Sadece tek bir taramayla bile yeni kümeler bulabilir. Fakat tarama sayısı artırılması durumunda küme kalitesi de artar.

Bu algoritma temelde 2 adımdan oluşmaktadır [52].

- Adım 1: Verinin doğal kümeleme yapısını korumaya çalışan kümeleme özellik ağacının başlangıç belleğini oluşturmak için veri tabanı taranır.
- Adım 2: Kümeleme özellik ağacının yaprak düğümlerini kümelemek için rastgele seçilmiş bir kümeleme algoritması kullanılır. Seyrek kümeler aykırı değerler olarak kaldırılır ve yoğun kümeler daha büyük kümeler halinde gruplandırılır.

Birch hiyerarşik kümeleme algoritması, diğer hiyerarşik kümeleme modellerine göre ölçeklenebilirdir ve bir önceki adıma geri dönülebilir.

3.3.3.3. Gauss Karma Model Kümeleme

Gauss karma model kümelemesi hızlı bir model olmayabilir fakat bazı durumlarda fazlasıyla etkili bir modeldir. Modelin başlangıcında her veri noktası kendi kümesidir. Daha sonra hangi çiftin birbirine en yakın (tek bağlantı) veya en uzak (tam bağlantı) olduğunu belirleyen kümeler üzerinde yinelenir. Bu kümeler daha sonra birleştirilerek yeni bir küme oluşturulur. Model üzerinde herhangi bir durdurma kriteri uygulanmazsa algoritma sonunda tüm kümeleri birleştirilerek tek bir küme haline getirir. Tek bağlantı veya tam bağlantı yaklaşımlarından verilere en uygun olanı seçilmelidir. Gauss karma modellerde önemli problemlerden birisi veri setinin bölünmesine uygun küme sayısının belirlenmesidir [67].

3.3.4. Kümeleme Performans Ölçütleri

Kümeleme algoritmaları, verileri belli kümelere ayırsa da her birinin aldığı parametreler farklı olduğu için farklı küme yapılarının ortaya çıkması kaçınılmazdır ve hangisinin daha iyi sonuç verdiğinin gözle kontrolle anlaşılması zor olabilir. Bu yüzden kümelerin değerlendirilmesi ve model sonuçlarının kıyaslanmasında objektif ölçütlere ihtiyaç duyulduğu için bazı performans metrikleri kullanılmaktadır [68]. Model sonuçlarından

elde edilen kümeler modellerin performansları çalışma süresi, Calinski-Harabasz İndeksi ve Davies-Bouldin İndeksi ile kıyaslanacaktır. Bu performans metriklerinin sonuçlarına göre şirket yetkililerin değerlendirmeleri sonucunda uygun model ve model sonuçları tezin devamında yer alan müşteri ilişkileri geliştirme stratejileri bölümünde kullanılacaktır.

4. UYGULAMA

Bu bölümde ilk olarak metodolojiden bahsedilmiştir. Sonrasında veri analizi ve verinin kümelemeye hazırlanması, öznitelik seçimi, küme sayısının belirlenmesi, kümeleme analizi, metotlarının sonuçlarının karşılaştırılması ve segmentlere göre müşteri ilişkileri geliştirme stratejileri başlıklarına yer verilmiştir.

4.1. Metodoloji

Araştırmanın genel yaklaşımı beş adımdan oluşmaktadır. Birincisi veri önışlemedir ve bu kısımda veri analizi yapılarak veri setinin kümeleme hazır hale getirilmesi amaçlanmaktadır. İkinci adım, eldeki öznitelikler içinden literatüre de bağılı kalarak öznitelik seçimi yapmaktır. Üçüncüsü, belirlenen özniteliklerle kaç adet küme ile çalışılması gerektiğine karar vermektir. Dördüncüsü, belirlenen öznitelikler ve küme sayısı ile üç kümeleme metodunu denemek, karşılaştırmak ve sektöre ve müşteri verilerine en uygun kümeleme metodunun seçmektir. Sonuncusu ise seçilen metodun sonuçlarına göre müşteri ilişkileri geliştirme önerilerinde bulunmaktır.

4.2. Veri Analizi ve Verinin Kümelemeye Hazırlanması

Bu bölümde tez için kullanılan verilerin analizi yapılmıştır ve analiz işlemi için Python programlama dili kullanılmıştır.

4.2.1. Genel Veri Tanıtımı

Toplamda 547.227 satır ve 15 sütundan oluşan bir veri seti bulunmaktadır. Her satır bir kullanıcının verisini içermektedir. İlk olarak ne kadar eksik veri olduğuna ve veri tiplerine bakılmıştır ve Çizelge 4.1’de bu bilgiler verilmiştir.

Çizelge 4.1. Veri Sütunları, Veri Sayıları ve Veri Tipleri

| <i>Veri Sütunları</i> | <i>Veri Sayısı</i> | <i>Veri Tipi</i> |
|---|--------------------|------------------|
| Şehir | 547.227 | Nitel |
| Müşteri Numarası | 547.227 | Nitel |
| İlk Sipariş Tarihi | 547.227 | Nitel |
| Son Sipariş Tarihi | 547.227 | Nitel |
| Toplam Sipariş Sayısı | 547.227 | Nicel |
| Başarılı Sipariş Sayısı | 547.227 | Nicel |
| İptal Sipariş Sayısı | 547.227 | Nicel |
| İptal Oranı | 547.227 | Nicel |
| Toplam Sepet Tutarı | 547.227 | Nicel |
| Kuponlu Sipariş Sayısı | 547.227 | Nicel |
| Kampanyalı Sipariş Sayısı | 547.227 | Nicel |
| Kuponlu Verilen Siparişlerin Toplam Sepet Tutarı | 352.497 | Nicel |
| Kampanyalı Verilen Siparişlerin Toplam Sepet Tutarı | 216.167 | Nicel |
| Kuponlu Sipariş Tutarı | 352.496 | Nicel |
| Kampanyalı Sipariş Tutarı | 216.161 | Nicel |

Veri setinin 11 özniteliği nicel, 4 tanesi nitel verilerden oluşmaktadır. Birden fazla geçen müşteri numarasının olup olmadığı incelendiğinde müşteri numarasının her müşteriye özel olduğu gözlemlenip veriye katkısı olmayacağı için silinmiştir.

Boş olmayan Kuponlu Verilen Siparişlerin Toplam Sepet Tutarı ve Kampanyalı Verilen Siparişlerin Toplam Sepet Tutarı değerlerinden sıfır olan olup olmadığı incelendiğinde sıfır olan değer gözlemlenmemiştir ve boş değerler yerine sıfır atanmıştır.

Şirketin kupon tanımlamalarında 5 ve katları dışında kupon tanımlaması olmadığı için kalan verilerin şirket testleri için kullanılan hesaplar olduğu düşünülüp bu satılar veri setinden çıkarılmıştır. Bu haliyle veri sayısı 544.772 adet kalmıştır.

4.2.2. Nümerik Özniteliklerin İstatiksel Verileri

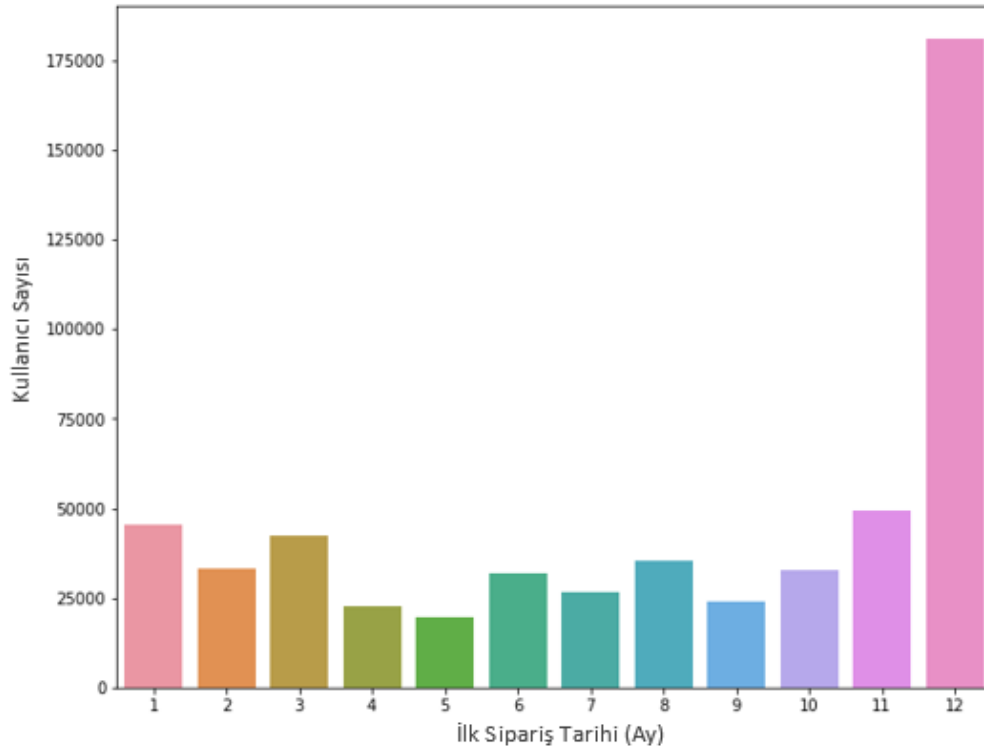
Çizelge 4.2. Nümerik Özniteliklerin İstatistiksel Veri Tablosu

| | <i>Toplam Sipariş Sayısı</i> | <i>Başarılı Sipariş Sayısı</i> | <i>İptal Sipariş Sayısı</i> | <i>İptal Oranı</i> | <i>Toplam Sepet Tutarı</i> | <i>Kuponlu Sipariş Sayısı</i> | <i>Kampanyalı Sipariş Sayısı</i> | <i>Kuponlu Verilen Siparişlerin Toplam Sepet Tutarı</i> | <i>Kampanyalı Verilen Siparişlerin Toplam Sepet Tutarı</i> | <i>Kuponlu Sipariş Tutarı</i> | <i>Kampanyalı Sipariş Tutarı</i> |
|-----------------------|------------------------------|--------------------------------|-----------------------------|--------------------|----------------------------|-------------------------------|----------------------------------|---|--|-------------------------------|----------------------------------|
| Adet | 544.772 | 544.772 | 544.772 | 544.772 | 544.772 | 544.772 | 544.772 | 544.772 | 544.772 | 544.772 | 544.772 |
| Ortalama | 2,1369 | 1,9809 | 0,1560 | 0,0420 | 220,4116 | 0,9541 | 0,5592 | 105,7532 | 66,9901 | 35,3195 | 11,9742 |
| Standart Sapma | 2,9166 | 2,6856 | 0,5208 | 0,1271 | 408,9687 | 1,0374 | 1,0125 | 116,3629 | 146,7635 | 37,6915 | 23,5141 |
| Minimum | 1,0 | 1,0 | 0,0 | 0,0 | -8,0 | 0,0 | 0,0 | 0,0 | 0,0 | 0,0 | 0,0 |
| 25% | 1,0 | 1,0 | 0,0 | 0,0 | 90,3 | 0,0 | 0,0 | 0,0 | 0,0 | 0,0 | 0,0 |
| 50% | 1,0 | 1,0 | 0,0 | 0,0 | 122,7 | 1,0 | 0,0 | 96,1 | 0,0 | 40,0 | 0,0 |
| 75% | 2,0 | 2,0 | 0,0 | 0,0 | 224,4 | 1,0 | 1,0 | 144,2 | 102,8 | 40,0 | 20,0 |
| Maksimum | 214,0 | 198,0 | 62,0 | 1,0 | 49,5 | 66,0 | 63,0 | 3.598,3 | 19.823,6 | 1.830,0 | 1.960,0 |

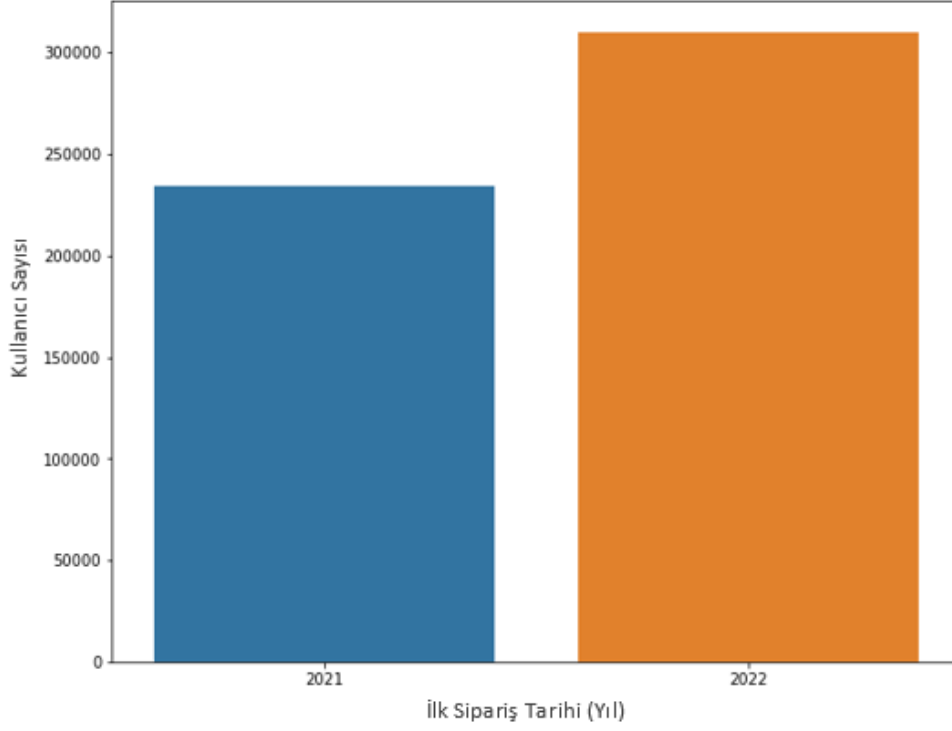
Çizelge 4.2 incelendiğinde siparişlerin yüzde 50'sinden fazlasının 1 adet sipariş vermiş olan müşterilerden oluştuğu tespit edilmiştir.

Toplam Sipariş Sayısı ve Başarılı Sipariş Sayısı için maksimum değeri alan siparişe bakıldığında değerlerin makul olduğu görülmüştür. Toplam Sepet Tutarı ortalaması değeri 20 TL'den az olan siparişler mümkün olmadığı için bu siparişler de veri setinden çıkarılmıştır. Sipariş sayısı 10'un üzerinde ve iptal oranı %80 üzerinde olan siparişlerin de test verisi olabileceği düşünülerek veri setinden çıkarılmışlardır. Sonuçta 544.393 adet veri kalmıştır.

4.2.3. Sipariş Verilerine Göre Analiz



Şekil 4.1. İlk Sipariş Tarihinin Aylara Göre Dağılımı



Şekil 4.2. İlk Sipariş Tarihinin Yıllara Göre Dağılımı

Şekil 4.1 ve Şekil 4.2 incelenerek veri seti analiz edildiğinde 2021 yılının 12. ayında veri yığılması olduğu fark edilmiştir. Bunun sebebi araştırıldığında 2021 yılı 12. ayında şirket tarafından agresif bir kupon çalışması yapıldığı ve bu sebeple ilk siparişini veren kullanıcıların bu kısma yığıldığı tespit edilmiştir. Bu yığılmanın ileri aşamalarda yapılacak sınıflandırma algoritmalarını yanıltma ihtimali düşünüldüğünden 2021 yılında ilk siparişini veren kullanıcılar veri setinden çıkarılmıştır. 2022 yılı verileriyle çalışmalara devam edilmiştir. Son durumda 310.096 satır 14 sütun verimiz kalmıştır ve verinin son durumu Çizelge 4.3’de verilmiştir.

Çizelge 4.3. Veri Temizliği Sonunda Kalan Veri Seti

| <i>Veri Sütunları</i> | <i>Veri Sayısı</i> | <i>Veri Tipi</i> |
|---|--------------------|------------------|
| Şehir | 310.096 | Nitel |
| İlk Sipariş Tarihi | 310.096 | Nitel |
| Son Sipariş Tarihi | 310.096 | Nitel |
| Toplam Sipariş Sayısı | 310.096 | Nicel |
| Başarılı Sipariş Sayısı | 310.096 | Nicel |
| İptal Sipariş Sayısı | 310.096 | Nicel |
| İptal Oranı | 310.096 | Nicel |
| Toplam Sepet Tutarı | 310.096 | Nicel |
| Kuponlu Sipariş Sayısı | 310.096 | Nicel |
| Kampanyalı Sipariş Sayısı | 310.096 | Nicel |
| Kuponlu Verilen Siparişlerin Toplam Sepet Tutarı | 310.096 | Nicel |
| Kampanyalı Verilen Siparişlerin Toplam Sepet Tutarı | 310.096 | Nicel |
| Kuponlu Sipariş Tutarı | 310.096 | Nicel |
| Kampanyalı Sipariş Tutarı | 310.096 | Nicel |

4.2.4. Verinin Kümelemeye Uygun Hale Getirilmesi

Büyük veri içerisindeki öznitelikler analize hazır hale getirilmek için son düzenlemeye geçilmiştir. Tez çalışması kapsamında RFM kriterleri kullanılacağı için son sipariş tarihi ile günümüz arasındaki fark alınarak güncellik özneliği oluşturulmuştur. Toplam sipariş sayısı verisi sıklık kriteri için, toplam sepet tutarı verisi ise para kriteri için kullanılacaktır.

Yine tez çalışmamız kapsamında RFM kriterlerine ek olarak bir kriter daha seçilmek istenmiştir. Bu sebeple ilk sipariş tarihi ile son sipariş tarihi arasındaki fark alınarak literatür bölümünde de bahsi geçen uzunluk (Length) kriteri oluşturulmuştur. Bu kriter “Gün farkı” olarak adlandırılmıştır. Kupon ve kampanya verileri de incelenmek istendiği için eldeki verilerden “Kuponlu Sipariş Sayısı Yüzdesi”, “Kuponlu Sipariş Tutarı Yüzdesi”, “Kampanyalı Sipariş Sayısı Yüzdesi” ve “Kampanyalı Sipariş Tutarı Yüzdesi” kriterleri oluşturulmuştur. Son durumda öznitelik seçimi bölümüne aktarılan 8 adet öznitelik vardır ve bunlar aşağıda tanımlarıyla birlikte listelenmiştir.

- Güncellik, baz alınan zaman diliminin son tarihi ile bu zaman diliminde verilen son sipariş tarih arasındaki fark alınarak hesaplanmaktadır.

- Sıklık, baz alınan zaman dilimi içerisinde verilen siparişlerin sayısal olarak toplamıdır.
- Para, baz alınan zaman dilimi içerisinde verilen siparişlerin parasal değerlerinin toplamıdır.
- Gün Farkı (L), baz alınan zaman dilimi içerisinde ilk sipariş tarihi ile son sipariş tarihi arasındaki farktır.
- Kuponlu Sipariş Sayısı Yüzdesi, baz alınan zaman dilimi içerisinde kupon kullanılan siparişlerin toplam siparişler içerisindeki yüzdesidir.
- Kuponlu Sipariş Tutarı Yüzdesi, baz alınan zaman dilimi içerisinde kupon kullanılan siparişlerin sepet tutarının toplam siparişlerin sepet tutarı içerisindeki yüzdesidir.
- Kampanyalı Sipariş Sayısı Yüzdesi, baz alınan zaman dilimi içerisinde kampanya kullanılan siparişlerin toplam siparişler içerisindeki yüzdesidir.
- Kampanyalı Sipariş Tutarı Yüzdesi, baz alınan zaman dilimi içerisinde kampanya kullanılan siparişlerin sepet tutarının toplam siparişlerin sepet tutarı içerisindeki yüzdesidir.

Şehir özniteliğinin bu kümeye dahil edilmemesinin sebebi bu verinin kullanıcının son sipariş verdiği tarihi ifade etmesidir. Yani kullanıcının yaşadığı şehri ifade etmemektedir. Bu yüzden şehir özniteliğine göre segment çalışması yapıldığı takdirde herhangi bir sebepten dolayı başka şehirden sipariş vermiş fakat o şehirde yaşamayan kullanıcıları segment dışında bırakmış olma riskimiz vardır. Bu sebeplerden dolayı öznitelik seçimi analizinde bu özniteliğe yer verilmeyecektir.

4.3. Öznitelik Seçimi

Çalışmamız kapsamında hem literatür taraması sonrası hem de şirket için de önem arz etmesi sebebiyle RFM kriterleri kullanılacaktır. Bunun yanı sıra tezin hedeflerinden biri, hızlı tüketim ürünleri sektöründe online alışveriş yapan müşterilerin segmentasyonu için kullanılacak uygun kriterlerin ortaya konması olduğu için öznitelik seçimi yapılarak bir adet kriter daha seçmeye karar verilmiştir. Ek olarak bir tane özniteliğin seçilmesinin

sebebi eldeki veri setimizin çok fazla veriden oluşması ve veri boyutunun daha fazla artırılmasının problemlerin çözümünü zorlaştıracak olmasıdır.

RFM'e ek olarak kullanılacak özniteliğin seçimi için Calinski-Harabasz İndeksi ve Davies-Bouldin İndeksi kullanılarak belirlenen üç kümeleme metodu için de 2 kümeden 10 kümeye kadar deneme yapılmıştır.

4.3.1. Calinski-Harabasz İndeksi'ne göre Öznitelik Seçimi

- *K-Ortalamalar Kümeleme ile Deneme:*

Çizelge 4.4. Calinski-Harabasz İndeksi'ne Göre K-Ortalamalar Deneme Skorları

| <i>Küme Sayısı</i> | <i>İlk Seçilen Öznitelik</i> | <i>Skoru</i> |
|--------------------|--------------------------------|--------------|
| 2 | Kuponlu Sipariş Sayısı Yüzdesi | 3876482,60 |
| 3 | Kuponlu Sipariş Sayısı Yüzdesi | 17667182,57 |
| 4 | Kuponlu Sipariş Sayısı Yüzdesi | 31063142,42 |
| 5 | Kuponlu Sipariş Sayısı Yüzdesi | 59305241,36 |
| 6 | Kuponlu Sipariş Sayısı Yüzdesi | 101980025,71 |
| 7 | Kuponlu Sipariş Sayısı Yüzdesi | 127940624,59 |
| 8 | Kuponlu Sipariş Sayısı Yüzdesi | 212952495,58 |
| 9 | Kuponlu Sipariş Sayısı Yüzdesi | 256710768,72 |
| 10 | Kuponlu Sipariş Sayısı Yüzdesi | 318254674,78 |

Çizelge 4.4'de yer alan Calinski-Harabasz İndeksi sonuçlarına göre K-Ortalamalar metodu ile denenen tüm kümelere göre "Kuponlu Sipariş Sayısı Yüzdesi" seçilen öznitelik olmuştur.

- *Gauss Karma Model ile Deneme:*

Çizelge 4.5. Calinski-Harabasz İndeksi'ne Göre Gauss Karma Model Deneme Skorları

| <i>Küme Sayısı</i> | <i>İlk Seçilen Öznitelik</i> | <i>Skoru</i> |
|--------------------|--------------------------------|--------------|
| 2 | Kuponlu Sipariş Sayısı Yüzdesi | 2843671,40 |
| 3 | Kuponlu Sipariş Sayısı Yüzdesi | 10280450,46 |
| 4 | Kuponlu Sipariş Sayısı Yüzdesi | 24012487,66 |
| 5 | Kuponlu Sipariş Sayısı Yüzdesi | 37958800,72 |
| 6 | Kuponlu Sipariş Sayısı Yüzdesi | 50167922,49 |
| 7 | Kuponlu Sipariş Sayısı Yüzdesi | 50656173,03 |
| 8 | Kuponlu Sipariş Sayısı Yüzdesi | 181632568,70 |
| 9 | Kuponlu Sipariş Sayısı Yüzdesi | 176670789,17 |
| 10 | Kuponlu Sipariş Sayısı Yüzdesi | 283974588,35 |

Çizelge 4.5’de yer alan Calinski-Harabasz İndeksi sonuçlarına göre Gauss Karma Model metodu ile denenen tüm kümelere göre “Kuponlu Sipariş Sayısı Yüzdesi” seçilen öznitelik olmuştur.

- *Birch Hiyerarşik Kümeleme ile Deneme:*

Çizelge 4.6. Calinski-Harabasz İndeksi'ne Göre Birch Hiyerarşik Kümeleme Deneme Skorları

| <i>Küme Sayısı</i> | <i>İlk Seçilen Öznitelik</i> | <i>Skoru</i> |
|--------------------|--------------------------------|--------------|
| 2 | Kuponlu Sipariş Sayısı Yüzdesi | 3212758,27 |
| 3 | Kuponlu Sipariş Sayısı Yüzdesi | 17667182,58 |
| 4 | Kuponlu Sipariş Sayısı Yüzdesi | 28473589,55 |
| 5 | Kuponlu Sipariş Sayısı Yüzdesi | 37567843,42 |
| 6 | Kuponlu Sipariş Sayısı Yüzdesi | 47847562,98 |
| 7 | Kuponlu Sipariş Sayısı Yüzdesi | 48667182,58 |
| 8 | Kuponlu Sipariş Sayısı Yüzdesi | 162983653,73 |
| 9 | Kuponlu Sipariş Sayısı Yüzdesi | 156735947,63 |
| 10 | Kuponlu Sipariş Sayısı Yüzdesi | 238892354,90 |

Çizelge 4.6’da yer alan Calinski-Harabasz İndeksi sonuçlarına göre Birch Hiyerarşik Kümeleme Model metodu ile denenen tüm kümelere göre “Kuponlu Sipariş Sayısı Yüzdesi” seçilen öznitelik olmuştur.

4.3.2. Davies-Bouldin İndeksi'ne Göre Öznitelik Seçimi

- *K-Ortalamalar Kümeleme ile Deneme:*

Çizelge 4.7. Davies-Bouldin İndeksi'ne Göre Göre K-Ortalamalar Deneme Skorları

| <i>Küme Sayısı</i> | <i>Seçilen Öznitelik</i> | <i>Skoru</i> |
|--------------------|-----------------------------------|--------------|
| 2 | Kuponlu Sipariş Sayısı Yüzdesi | 0,1432 |
| 3 | Kuponlu Sipariş Sayısı Yüzdesi | 0,2048 |
| 4 | Kampanyalı Sipariş Sayısı Yüzdesi | 0,3259 |
| 5 | Kuponlu Sipariş Sayısı Yüzdesi | 0,2478 |
| 6 | Kuponlu Sipariş Sayısı Yüzdesi | 0,2654 |
| 7 | Kuponlu Sipariş Sayısı Yüzdesi | 0,2854 |
| 8 | Kuponlu Sipariş Sayısı Yüzdesi | 0,3210 |
| 9 | Kuponlu Sipariş Sayısı Yüzdesi | 0,3083 |
| 10 | Kuponlu Sipariş Sayısı Yüzdesi | 0,2847 |

Çizelge 4.7'de yer alan Davies-Bouldin İndeksi sonuçlarına göre küme sayısının 4 olduğu durumda “Kampanyalı Sipariş Sayısı Yüzdesi” özniteliği seçilmiş olmasına rağmen diğer küme denemelerinde “Kuponlu Sipariş Sayısı Yüzdesi” özniteliği seçilen öznitelik olmuştur.

- *Gauss Karma Model ile Deneme:*

Çizelge 4.8. Davies-Bouldin İndeksi'ne Göre Gauss Karma Model Deneme Skorları

| <i>Küme Sayısı</i> | <i>İlk Seçilen Öznitelik</i> | <i>Skoru</i> |
|--------------------|--------------------------------|--------------|
| 2 | Kuponlu Sipariş Sayısı Yüzdesi | 0,1422 |
| 3 | Kuponlu Sipariş Sayısı Yüzdesi | 0,2891 |
| 4 | Kuponlu Sipariş Sayısı Yüzdesi | 0,3603 |
| 5 | Kuponlu Sipariş Sayısı Yüzdesi | 0,2792 |
| 6 | Kuponlu Sipariş Sayısı Yüzdesi | 0,3447 |
| 7 | Kuponlu Sipariş Sayısı Yüzdesi | 0,3008 |
| 8 | Kuponlu Sipariş Sayısı Yüzdesi | 0,3454 |
| 9 | Kuponlu Sipariş Sayısı Yüzdesi | 0,3063 |
| 10 | Kuponlu Sipariş Sayısı Yüzdesi | 0,2854 |

Çizelge 4.8’de yer alan Davies-Bouldin İndeksi sonuçlarına göre Gauss Karma Model metodu ile denenen tüm kümelere göre “Kuponlu Sipariş Sayısı Yüzdesi” seçilen öznitelik olmuştur.

- *Birch Hiyerarşik Kümeleme ile Deneme:*

Çizelge 4.9. Davies-Bouldin İndeksi’ne Göre Birch Hiyerarşik Kümeleme Deneme Skorları

| <i>Küme Sayısı</i> | <i>İlk Seçilen Öznitelik</i> | <i>Skoru</i> |
|--------------------|-----------------------------------|--------------|
| 2 | Kampanyalı Sipariş Sayısı Yüzdesi | 0,0918 |
| 3 | Kuponlu Sipariş Sayısı Yüzdesi | 0,2048 |
| 4 | Kuponlu Sipariş Sayısı Yüzdesi | 0,2384 |
| 5 | Kuponlu Sipariş Sayısı Yüzdesi | 0,3864 |
| 6 | Kuponlu Sipariş Sayısı Yüzdesi | 0,3728 |
| 7 | Kuponlu Sipariş Sayısı Yüzdesi | 0,3758 |
| 8 | Kuponlu Sipariş Sayısı Yüzdesi | 0,3462 |
| 9 | Kuponlu Sipariş Sayısı Yüzdesi | 0,3173 |
| 10 | Kuponlu Sipariş Sayısı Yüzdesi | 0,2973 |

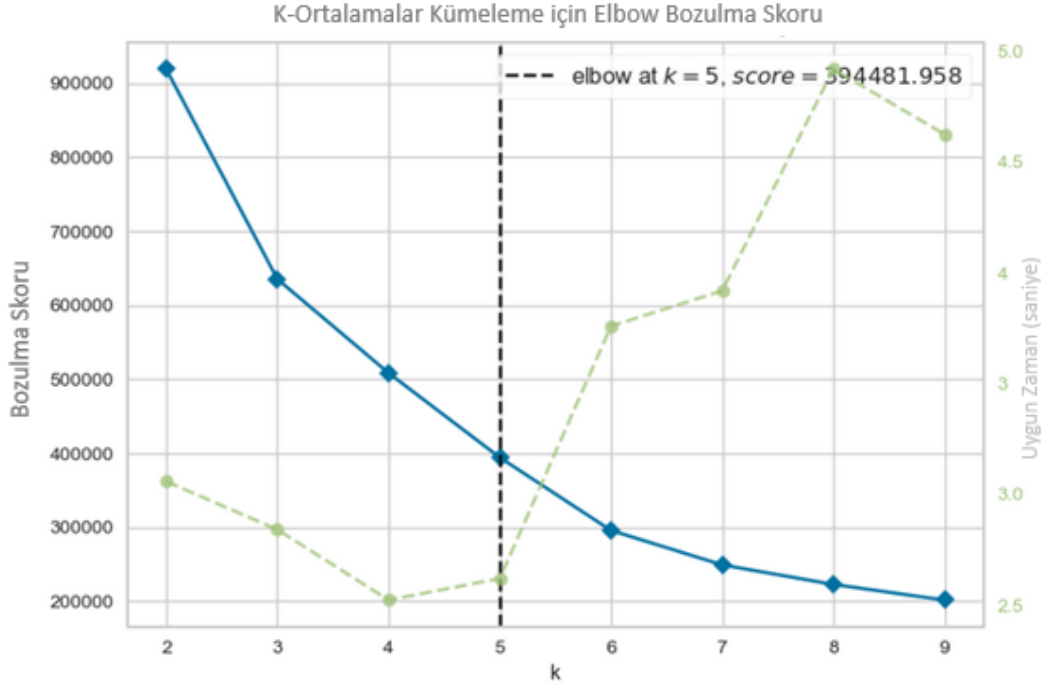
Çizelge 4.9’da yer alan Davies-Bouldin İndeksi sonuçlarına göre Birch Hiyerarşik Kümeleme metodu ile denemelerde küme sayısının 2 olduğu durumda “Kampanyalı Sipariş Sayısı Yüzdesi” özniteliği seçilmiş olmasına rağmen diğer küme denemelerinde “Kuponlu Sipariş Sayısı Yüzdesi” özniteliği seçilen öznitelik olmuştur.

İki metodun da üç model için de denemelerde büyük çoğunlukla “Kuponlu Sipariş Sayısı Yüzdesi” özniteliğini seçtiği gözlemlenmiştir. Şirket yetkilileriyle yapılan görüşmede de kupon kullanım durumlarının yer alınan sektörde kıymetli bir veri olduğu bilgisi alınmıştır. Sonuçlar ve alınan görüşlerle birlikte dördüncü öznitelik olarak “Kuponlu Sipariş Sayısı Yüzdesi” özniteliğinin seçilmesine karar verilmiştir.

4.4. Küme Sayısının Belirlenmesi

Seçilen bu 4 öznitelik ile küme sayısı belirlemek için tüm yöntemlerde uygulanacak uygun ve ortak bir değer bulunmak istenmiştir. Bunun için hem Elbow grafiğine bakmak

hem de şirket yetkililerinin bilgisine başvurmak uygun görülmüştür. Elbow'un uygulanacağı yöntem olarak benzer problemler için literatürde en sık kullanılan K-Ortalamlar seçilmiştir. Şekil 4.3'de yer alan Elbow grafiğine bakıldığında küme sayısının 5 olması gerektiği gözlemlenmiştir.



Şekil 4.3. Elbow Metodu Grafiği

Elbow metodunda bozulma skoru (distortion) hesaplaması kullanılmıştır. Bozulma skoru ilgili kümelerin elemanlarının küme merkezlerinden uzaklıklarının karelerinin ortalaması alınarak hesaplanır. Grafikte bozulma skoru ile uygun zaman kıyaslanmıştır. Kıyaslama sonucunda 5 adet küme kullanımını önerilmiştir. Sonuçlar üzerinden şirket pazarlama yetkilileri ile yapılan görüşmelerden, müşteri devamlılığının sağlanabilmesi için 5 müşteri segmenti üzerinden işlem yapılmasının doğru olacağı bilgisi alınmıştır ve küme sayısı 5 olarak belirlenmiştir.

4.5. Kümeleme Analizi

Bu bölümde belirlenen küme sayısı ve özniteliklerle K- Ortalamalar, Gauss Karma Model ve Birch Hiyerarşik kümeleme metotlarıyla kümeleme analizi yapılacaktır.

4.5.1. K-Ortalamlar Kümeleme

K-Ortalamlar kümeleme modelinde iki parametreyle kümeleme denemeleri yapılmıştır. Bu parametreler “Lloyd” ve “Elkan” parametreleridir. Bu parametreler temelde küme merkezlerinin belirlenmesi için kullanılmaktadır. Lloyd ilkendirme yöntemi olarak önerilmiştir ve veri içerisinde rastgele bir alt küme üzerinde k parametresi için kümeleme çalıştırılarak elde edilir. Bu işlem n defa tekrarlanır ve $n*k$ adet son veri kullanılarak kümeleme merkezleri elde edilmiş olur. Elkan ise alt kümelerin daha doğru oluşturulabilmesi amacıyla kümeleme algoritması her çalıştırıldığında hangi merkezlerin kendilerini bölmeleri gerektiği konusunda yerel kararlar oluşturarak ilerlemektedir. Bayes’in bilgi kriteri (BIC) hesaplanarak bu bölünme kararları verilmektedir. Aşağıda bu iki parametreye göre sonuçların kıyaslandığı tablo yer almaktadır.

Çizelge 4.10’da görüldüğü üzere iki parametrenin skorları arasında belirgin fark oluşmamış durumdadır. Bu yüzden iki parametreden herhangi birinin sonuçları alınabilir durumdadır. Sürenin bu çalışmada çok önemli olmadığı değerlendirilerek diğer iki skorda daha iyi değerler alan Elkan parametresiyle yapılan kümeleme çalışması dikkate alınmıştır.

Çizelge 4.10. K-Ortalamlar Kümeleme Algoritma Kıyaslaması

| <i>Küme Sayısı</i> | <i>Algoritma</i> | <i>Calinski-Harabasz Skoru</i> | <i>Davies-Bouldin Skoru</i> | <i>Süre (sn)</i> |
|--------------------|------------------|--------------------------------|-----------------------------|------------------|
| 5 | Lloyd | 166234,2963 | 0,7745 | 3,0667 |
| 5 | Elkan | 166234,8589 | 0,7743 | 4,8087 |

Elkan parametresiyle yapılan K-Ortalamlar kümelemenin küme özellikleri aşağıda verilmiştir.

- Çizelge 4.11 incelendiğinde, 1. kümeye 20.278 kullanıcı yerleştiği görülmektedir. Bu kullanıcıların güncellikleri ortalama 53 gün, ortalama sipariş verme sayıları 7 ve toplam sepet tutarları 843 TL çıkmıştır. Kuponlu sipariş verme yüzdeleri ise ortalama %10’dur. Bu kullanıcıların güncellik durumlarının ortalamada düşük olmasına

rağmen kuponsuz sipariş verme eğiliminde oldukları ve sepet ortalamalarının yüksek olduğu tespit edilmiştir.

Çizelge 4.11. K-Ortalamlar Kümeleme, 1. Kümenin Özellikleri

| <i>1. kümenin Özellikleri</i> | <i>Güncellik (R)</i> | <i>Sıklık (F)</i> | <i>Para (M)</i> | <i>Kuponlu Sipariş Sayısı Yüzdesi</i> |
|-------------------------------|----------------------|-------------------|-----------------|---------------------------------------|
| Adet | 20278 | 20278 | 20278 | 20278 |
| Ortalama | 53,3770 | 6,8857 | 843,1225 | 0,1075 |
| Standart Sapma | 55,6012 | 3,2986 | 488,1603 | 0,1802 |
| Minimum | 0,0000 | 1,0000 | 102,0000 | 0,0000 |
| 25% | 11,0000 | 5,0000 | 510,0675 | 0,0000 |
| 50% | 31,0000 | 6,0000 | 711,0850 | 0,0000 |
| 75% | 81,0000 | 8,0000 | 1042,8700 | 0,1667 |
| Maksimum | 291,0000 | 28,0000 | 3986,0000 | 1,0000 |

- Çizelge 4.12 incelendiğinde, 2. kümeye 156.040 adet kullanıcı yerleştiği görülmektedir. Bu kullanıcıların güncellikleri ortalama 127 gün, ortalama sipariş verme sayıları 2 ve toplam sepet tutarları 141 TL çıkmıştır. Kuponlu sipariş verme yüzdeleri ise ortalama %2'dir. Bu kullanıcıların güncellik durumlarının ortalamada bir hayli düşük olduğu, kuponsuz sipariş verme eğiliminde oldukları ve sepet ortalamalarının düşük olduğu tespit edilmiştir.

Çizelge 4.12. K-Ortalamlar Kümeleme, 2. Kümenin Özellikleri

| <i>2. kümenin özellikleri</i> | <i>Güncellik (R)</i> | <i>Sıklık (F)</i> | <i>Para (M)</i> | <i>Kuponlu Sipariş Sayısı Yüzdesi</i> |
|-------------------------------|----------------------|-------------------|-----------------|---------------------------------------|
| Adet | 156040 | 156040 | 156040 | 156040 |
| Ortalama | 127,0794 | 1,5143 | 141,5292 | 0,0206 |
| Standart Sapma | 83,3614 | 0,8774 | 119,0084 | 0,0885 |
| Minimum | 0,0000 | 1,0000 | 20,0000 | 0,0000 |
| 25% | 53,0000 | 1,0000 | 56,2500 | 0,0000 |
| 50% | 126,0000 | 1,0000 | 98,9500 | 0,0000 |
| 75% | 191,0000 | 2,0000 | 183,0600 | 0,0000 |
| Maksimum | 299,0000 | 7,0000 | 1104,1500 | 0,5000 |

- Çizelge 4.13 incelendiğinde, 3. kümeye 1.363 adet kullanıcının yerleştiği görülmektedir. Bu kullanıcıların güncellikleri ortalama 29 gün, ortalama sipariş

verme sayıları 24 ve toplam sepet tutarları 3996 TL çıkmıştır. Kuponlu sipariş verme yüzdeleri ise ortalama %2'dir. Bu kullanıcıların güncellik durumları ortalama diğer iki kümeye göre daha iyi çıkmıştır. Daha sık ve güncel sipariş vericiler olup kuponsuz sipariş verme eğiliminde oldukları ve sepet ortalamalarının yüksek olduğu tespit edilmiştir.

Çizelge 4.13. K-Ortalamalar Kümeleme, 3. Kümenin Özellikleri

| 3. kümenin özellikleri | Güncellik (R) | Sıklık (F) | Para (M) | Kuponlu Sipariş Sayısı Yüzdesi |
|-------------------------------|----------------------|-------------------|-----------------|---------------------------------------|
| Adet | 1363 | 1363 | 1363 | 1363 |
| Ortalama | 29,4057 | 24,4263 | 3996,0827 | 0,0245 |
| Standart Sapma | 41,6569 | 12,2977 | 2373,5454 | 0,0515 |
| Minimum | 0,0000 | 3,0000 | 889,9100 | 0,0000 |
| 25% | 4,0000 | 17,0000 | 2596,8750 | 0,0000 |
| 50% | 12,0000 | 22,0000 | 3375,7800 | 0,0000 |
| 75% | 33,5000 | 29,0000 | 4558,7800 | 0,0400 |
| Maksimum | 263,0000 | 112,0000 | 22172,1600 | 0,8500 |

- Çizelge 4.14 incelendiğinde, 4. kümeye 66.921 adet kullanıcı yerleştiği görülmektedir. Bu kullanıcıların güncellikleri ortalama 62 gün, ortalama sipariş verme sayıları 1 ve toplam sepet tutarları 172 TL çıkmıştır. Kuponlu sipariş verme yüzdeleri ise ortalama %92'dir. Bu kullanıcıların güncellik durumlarının ortalama düşük ve kupon kullanarak bir defa düşük sepet tutarlı sipariş vermeye meyilli oldukları tespit edilmiştir.

Çizelge 4.14. K-Ortalamalar Kümeleme, 4. Kümenin Özellikleri

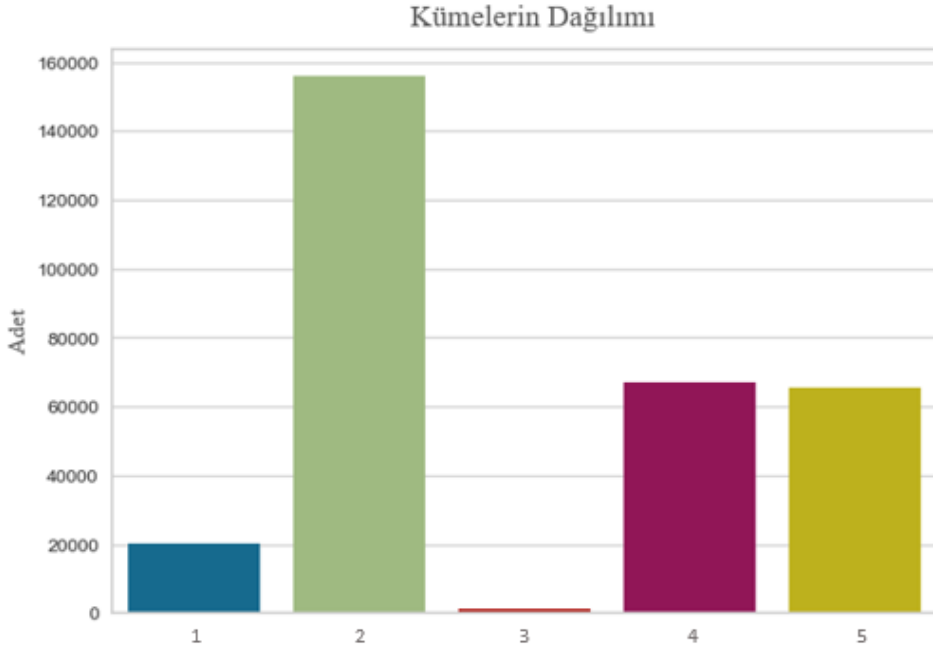
| 4. kümenin özellikleri | Güncellik (R) | Sıklık (F) | Para (M) | Kuponlu Sipariş Sayısı Yüzdesi |
|-------------------------------|----------------------|-------------------|-----------------|---------------------------------------|
| Adet | 66921 | 66921 | 66921 | 66921 |
| Ortalama | 62,6678 | 1,4216 | 172,1386 | 0,9158 |
| Standart Sapma | 40,8825 | 0,7577 | 92,5878 | 0,1849 |
| Minimum | 0,0000 | 1,0000 | 25,0000 | 0,3333 |
| 25% | 22,0000 | 1,0000 | 110,4000 | 1,0000 |
| 50% | 66,0000 | 1,0000 | 137,4800 | 1,0000 |
| 75% | 92,0000 | 2,0000 | 200,2500 | 1,0000 |
| Maksimum | 158,0000 | 7,0000 | 1250,7500 | 1,0000 |

- Çizelge 4.15 incelendiğinde, 5. kümeye 65.494 adet kullanıcı yerleştiği görülmektedir. Bu kullanıcıların güncellikleri ortalama 246 gün, ortalama sipariş verme sayıları 1 ve toplam sepet tutarları 141 TL çıkmıştır. Kuponlu sipariş verme yüzdeleri ise ortalama %95'dir. Bu kullanıcıların güncellik durumlarının ortalamada çok düşük ve kupon kullanarak bir defa düşük sepet ortalamalı sipariş vermeye meyilli oldukları tespit edilmiştir. 4. kümeye göre güncellik durumları daha kötü çıkmıştır.

Çizelge 4.15. K-Ortalamlar Kümeleme, 5. Kümenin Özellikleri

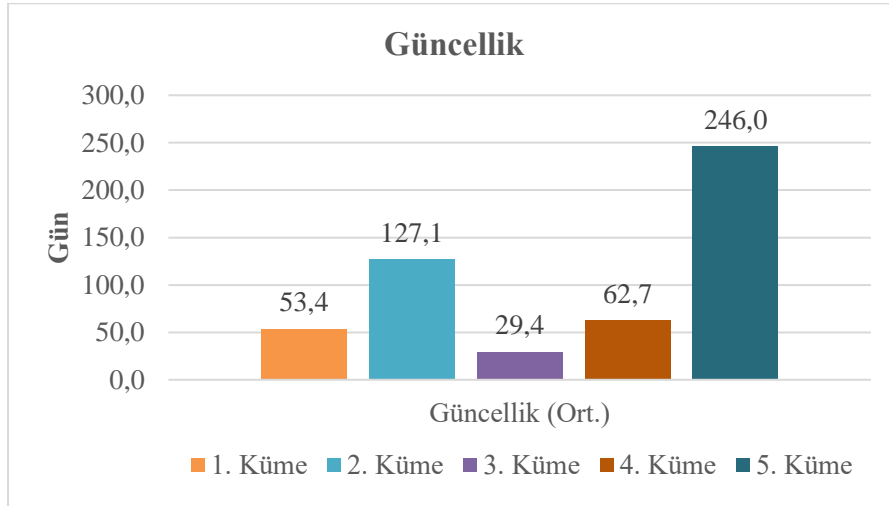
| <i>5. kümenin özellikleri</i> | <i>Güncellik (R)</i> | <i>Sıklık (F)</i> | <i>Para (M)</i> | <i>Kuponlu Sipariş Sayısı Yüzdesi</i> |
|-------------------------------|----------------------|-------------------|-----------------|---------------------------------------|
| Adet | 65494,0 | 65494,0 | 65494,0 | 65494,0 |
| Ortalama | 246,0462 | 1,4409 | 141,5645 | 0,9491 |
| Standart Sapma | 30,0921 | 0,8275 | 80,8055 | 0,1471 |
| Minimum | 153,0000 | 1,0000 | 20,2500 | 0,2000 |
| 25% | 223,0000 | 1,0000 | 94,2500 | 1,0000 |
| 50% | 240,0000 | 1,0000 | 106,7200 | 1,0000 |
| 75% | 278,0000 | 2,0000 | 160,2500 | 1,0000 |
| Maksimum | 299,0000 | 9,0000 | 1085,0000 | 1,0000 |

Son olarak Şekil 4.4'teki grafikte de kullanıcı dağılımları toplu olarak görülmektedir. 2. Kümeye yerleşen kullanıcıların yoğunluğu dikkat çekmektedir. 4. ve 5. kümeye yerleşen kullanıcı sayılarının dengeli olmasına rağmen 1. ve 3. kümeye yerleşen kullanıcı sayıları diğer kümelere göre belirgin şekilde düşük çıkmıştır.



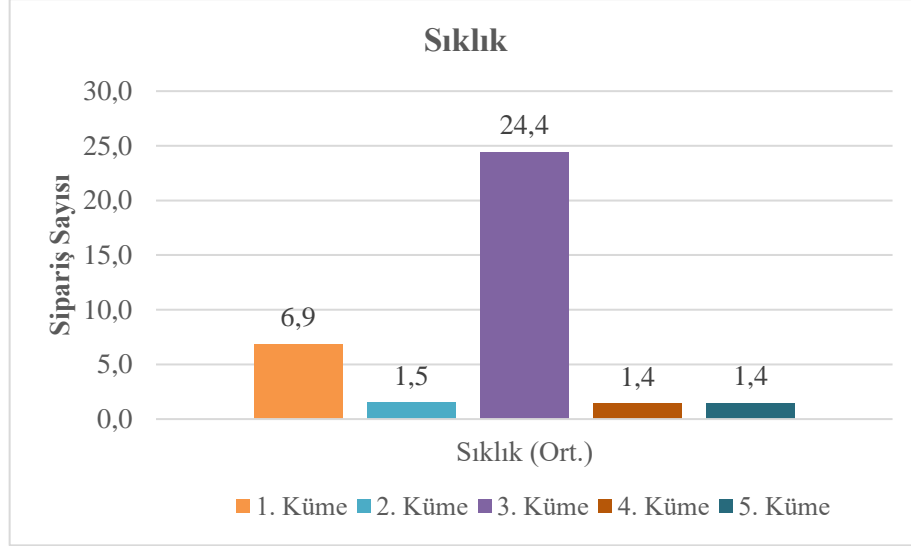
Şekil 4.4. K-Ortalamalar Kümeleme, Kullanıcı Sayılarının Kümelere Göre Dağılımları

Kriterlerin 5 küme için toplu grafikleri de aşağıdaki gibidir.



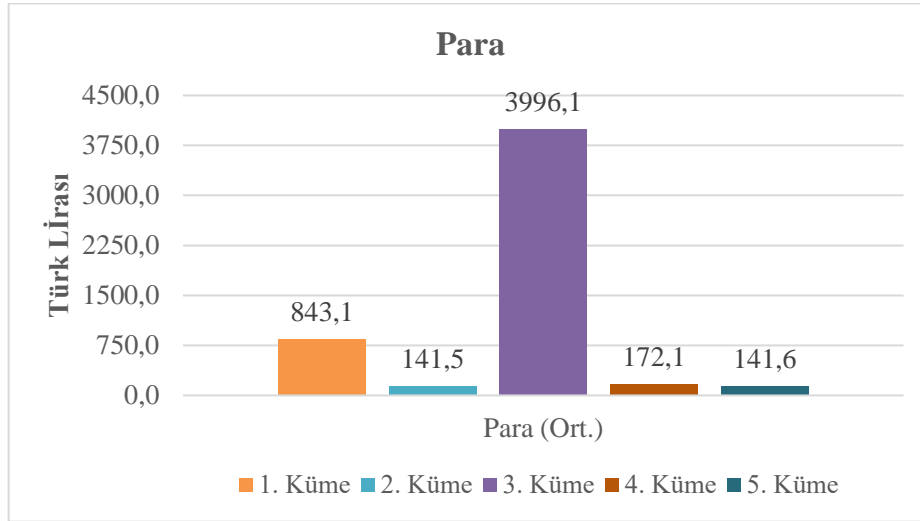
Şekil 4.5. K-Ortalamalar Kümeleme, Kümelere Göre Ortalama Güncellik Değerleri

Şekil 4.5'e göre, en güncel kullanıcılar 3. kümedeyken güncelliği en düşük olan kullanıcılar 5. kümede yer almıştır.



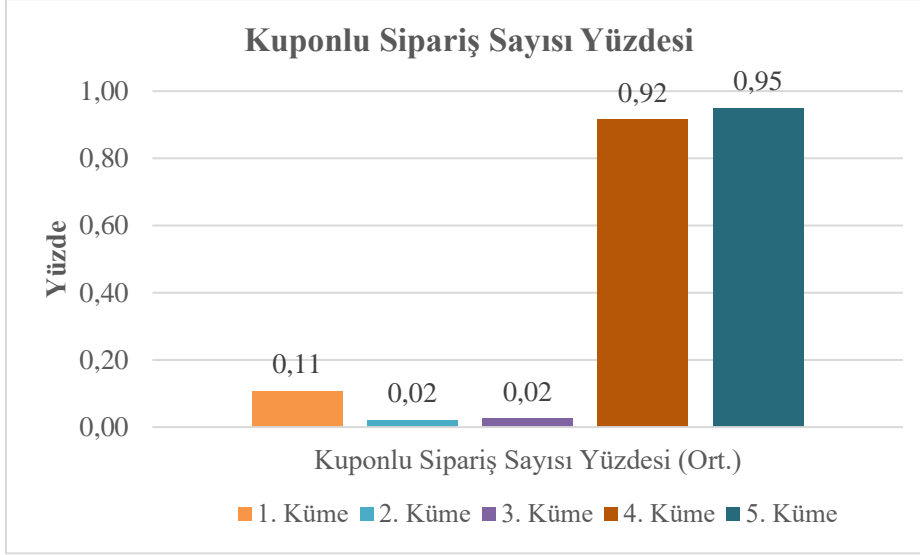
Şekil 4.6. K-Ortalamlar Kümeleme, Kümelere Göre Ortalama Sıklık Değerleri

Şekil 4.6'ya göre, en sık sipariş veren kullanıcılar 3. kümede yer alırken, en az sipariş veren kullanıcılar 4 ve 5. kümelerde yer almışlardır.



Şekil 4.7. K-Ortalamlar Kümeleme, Kümelere Göre Ortalama Para Değerleri

Şekil 4.7'ye göre, en yüksek sepet tutarlı kullanıcılar 3. kümede yer alırken en düşük sepet tutarlı kullanıcılar 2. ve 5. kümelerde yer almışlardır.



Şekil 4.8. K-Ortalamalar Kümeleme, Kümelere Göre Ortalama Kuponlu Sipariş Sayısı Yüzdesi Değerleri

Şekil 4.8'e göre, kupon kullanım yüzdesi en yüksek olan kullanıcılar 5. kümede yer alırken, en düşük olan kullanıcılar 2. ve 3. kümelerde yer almışlardır.

4.5.2. Birch Hiyerarşik Kümeleme

Hiyerarşik kümeleme yöntemlerinden Birch metodunun tercih edilmesinin sebebi büyük veri setlerinde hızlı çalışmasıdır. Veri sayısının büyüklüğünden kaynaklı diğer modellerin çalıştırılmasında sorun yaşanmıştır. Birch hiyerarşik kümeleme yöntemine göre çıkan sonuçlar aşağıda paylaşılmıştır. Birch metodunun elimizdeki veri setine ve özniteliklere göre başarılı bir kümeleme yapamadığı söylenebilir. Kriterlerin genel olarak istenilen düzeyde olduğu kullanıcıların 4 kümeye ayrıldığı kalanların tek bir kümeye yerleştirildiği tespit edilmiştir.

- Çizelge 4.16 incelendiğinde, 1. kümeye 47 adet kullanıcı yerleştiği görülmektedir. Bu kullanıcıların güncellikleri ortalama 26 gün, ortalama sipariş verme sayıları 42 ve toplam sepet tutarları 12.990 TL çıkmıştır. Kuponlu sipariş verme yüzdeleri ise ortalama %0,1'dir. Bu kullanıcıların güncellik durumları ortalamada yüksek ve kuponsuz sipariş verme eğiliminde oldukları ve sepet ortalamalarının çok yüksek olduğu tespit edilmiştir.

Çizelge 4.16. Birch Hiyerarşik Kümeleme, 1. Kümenin Özellikleri

| <i>1. kümenin özellikleri</i> | <i>Güncellik (R)</i> | <i>Sıklık (F)</i> | <i>Para (M)</i> | <i>Kuponlu Sipariş Sayısı Yüzdesi</i> |
|-------------------------------|----------------------|-------------------|-----------------|---------------------------------------|
| Adet | 47 | 47 | 47 | 47 |
| Ortalama | 26,9362 | 42,5319 | 12990,9071 | 0,0090 |
| Standart Sapma | 40,3282 | 24,0551 | 3430,6863 | 0,0251 |
| Minimum | 0,0000 | 6,0000 | 8632,7000 | 0,0000 |
| 25% | 1,5000 | 23,0000 | 10440,9250 | 0,0000 |
| 50% | 7,0000 | 39,0000 | 12011,7200 | 0,0000 |
| 75% | 38,0000 | 55,0000 | 14627,2850 | 0,0000 |
| Maksimum | 133,0000 | 112,0000 | 22172,1600 | 0,1429 |

- Çizelge 4.17 incelendiğinde, 2. kümeye 250 adet kullanıcı yerleştiği görülmektedir. Bu kullanıcıların güncellikleri ortalama 30 gün, ortalama sipariş verme sayıları 22 ve toplam sepet tutarları 5.718 TL çıkmıştır. Kuponlu sipariş verme yüzdeleri ise ortalama %0,2'dir. Bu kullanıcıların güncellik durumlarının ortalamada yüksek olup kuponsuz sipariş verme eğiliminde oldukları ve sepet ortalamalarının yüksek olduğu tespit edilmiştir. İlk kümeye göre sipariş verme sayıları ve ortalama toplam sepet tutarları daha düşük çıkmıştır.

Çizelge 4.17. Birch Hiyerarşik Kümeleme, 2. Kümenin Özellikleri

| <i>2. kümenin özellikleri</i> | <i>Güncellik (R)</i> | <i>Sıklık (F)</i> | <i>Para (M)</i> | <i>Kuponlu Sipariş Sayısı Yüzdesi</i> |
|-------------------------------|----------------------|-------------------|-----------------|---------------------------------------|
| Adet | 250 | 250 | 250 | 250 |
| Ortalama | 29,7240 | 21,5120 | 5718,2315 | 0,016325 |
| Standart Sapma | 41,3595 | 10,0582 | 1098,7769 | 0,039728 |
| Minimum | 0,0000 | 3,0000 | 4087,4600 | 0,000000 |
| 25% | 5,0000 | 14,2500 | 4830,2450 | 0,000000 |
| 50% | 13,0000 | 20,0000 | 5494,8950 | 0,000000 |
| 75% | 33,0000 | 28,0000 | 6423,1300 | 0,000000 |
| Maksimum | 200,0000 | 45,0000 | 9498,4300 | 0,285714 |

- Çizelge 4.18 incelendiğinde, 3. kümeye 309.166 adet kullanıcı yerleştiği görülmektedir. Bu kullanıcıların güncellikleri ortalama 133 gün, ortalama sipariş verme sayıları 2 ve toplam sepet tutarları 197 TL çıkmıştır. Kuponlu sipariş verme yüzdeleri ise ortalama %42'dir. Bu kullanıcıların güncellik durumlarının ortalamada

düşük, kuponsuz sipariş verme eğilimlerinin ortalama düzeyde olduğu ve ortalama toplam sepet tutarlarının düşük olduğu tespit edilmiştir.

Çizelge 4.18. Birch Hiyerarşik Kümeleme, 3. Kümenin Özellikleri

| 3. kümenin özellikleri | Güncellik (R) | Sıklık (F) | Para (M) | Kuponlu Sipariş Sayısı Yüzdesi |
|-------------------------------|----------------------|-------------------|-----------------|---------------------------------------|
| Adet | 309166 | 309166 | 309166 | 309166 |
| Ortalama | 133,3751 | 1,8519 | 197,9498 | 0,4168 |
| Standart Sapma | 92,1494 | 1,8690 | 257,0864 | 0,4666 |
| Minimum | 0,0000 | 1,0000 | 20,0000 | 0,0000 |
| 25% | 49,0000 | 1,0000 | 88,0000 | 0,0000 |
| 50% | 123,0000 | 1,0000 | 121,4000 | 0,0000 |
| 75% | 217,0000 | 2,0000 | 207,9400 | 1,0000 |
| Maksimum | 299,0000 | 26,0000 | 4424,0000 | 1,0000 |

- Çizelge 4.19 incelendiğinde, 4. kümeye 68 adet kullanıcı yerleştiği görülmektedir. Bu kullanıcıların güncellikleri ortalama 21 gün, ortalama sipariş verme sayıları 55 ve toplam sepet tutarları 6.560 TL çıkmıştır. Kuponlu sipariş verme yüzdeleri ise ortalama %0,1'dir. Bu kullanıcıların güncellik durumlarının ortalamada yüksek, kuponsuz sipariş verme eğilimlerinin yüksek ve ortalama toplam sepet tutarlarının yüksek olduğu tespit edilmiştir. 1. kümeden sonra ortalama toplam sepet tutarı en yüksek olan kümedir.

Çizelge 4.19. Birch Hiyerarşik Kümeleme, 4. Kümenin Özellikleri

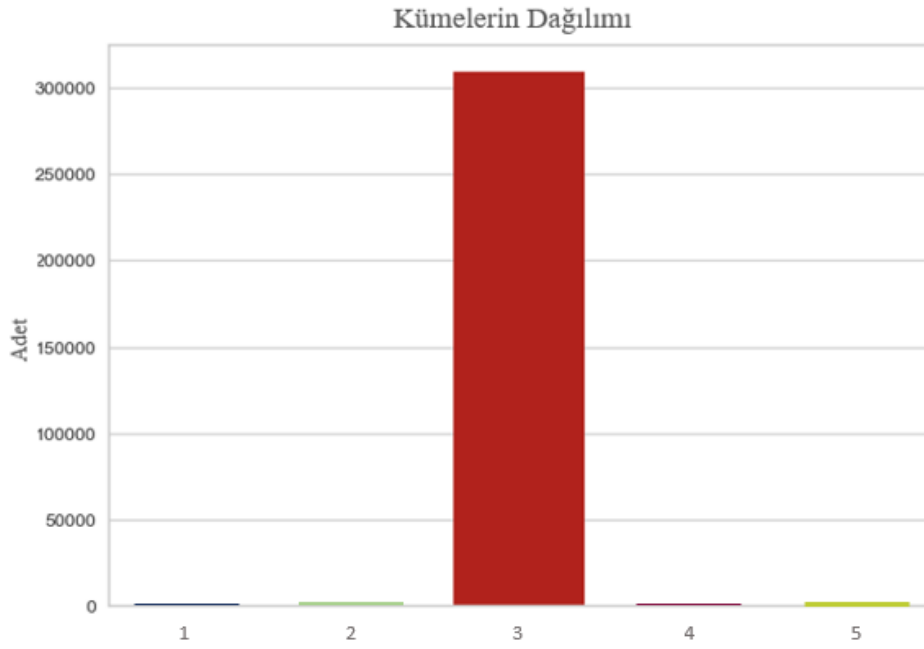
| 4. kümenin özellikleri | Güncellik (R) | Sıklık (F) | Para (M) | Kuponlu Sipariş Sayısı Yüzdesi |
|-------------------------------|----------------------|-------------------|-----------------|---------------------------------------|
| Adet | 68 | 68 | 68 | 68 |
| Ortalama | 21,5000 | 54,9559 | 6560,0620 | 0,0143 |
| Standart Sapma | 34,8080 | 11,6868 | 1790,7333 | 0,0234 |
| Minimum | 0,0000 | 40,0000 | 3581,8900 | 0,0000 |
| 25% | 3,0000 | 46,0000 | 4958,7200 | 0,0000 |
| 50% | 6,0000 | 51,5000 | 6193,9800 | 0,0000 |
| 75% | 17,2500 | 62,2500 | 7996,7973 | 0,0204 |
| Maksimum | 140,0000 | 96,0000 | 10532,8200 | 0,1020 |

- Çizelge 4.20 incelendiğinde, 5. kümeye 565 adet kullanıcı yerleştiği görülmektedir. Bu kullanıcıların güncellikleri ortalama 27 gün, ortalama sipariş verme sayıları 27 ve toplam sepet tutarları 6.560 TL çıkmıştır. Kuponlu sipariş verme yüzdeleri ise ortalama %0,2'dir. Bu kullanıcıların güncellik durumlarının ortalamada yüksek, kuponlu sipariş verme eğilimlerinin yüksek ve ortalama toplam sepet tutarlarının yüksek olduğu tespit edilmiştir.

Çizelge 4.20. Birch Hiyerarşik Kümeleme, 5. Kümenin Özellikleri

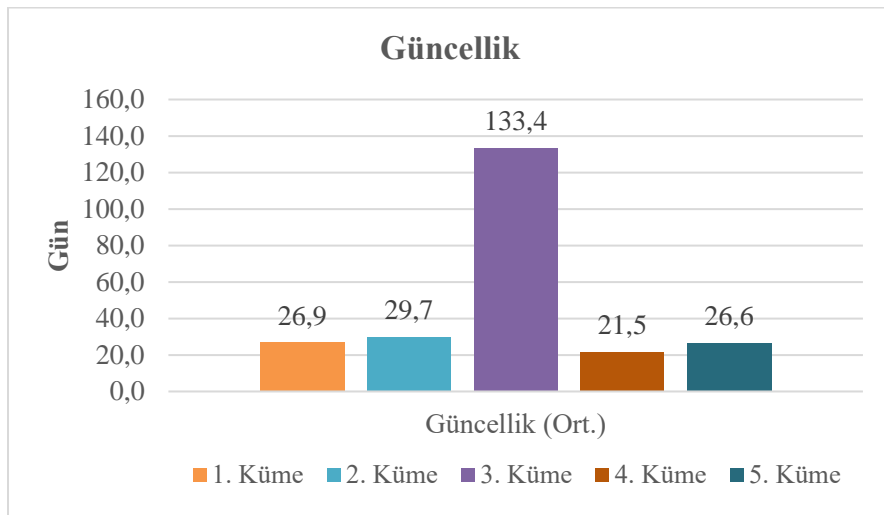
| <i>5. kümenin özellikleri</i> | <i>Güncellik (R)</i> | <i>Sıklık (F)</i> | <i>Para (M)</i> | <i>Kuponlu Sipariş Sayısı Yüzdesi</i> |
|-------------------------------|----------------------|-------------------|-----------------|---------------------------------------|
| Adet | 565 | 565,0000 | 565 | 565 |
| Ortalama | 26,6407 | 26,6230 | 3068,0581 | 0,0243 |
| Standart Sapma | 39,2042 | 6,7751 | 800,4756 | 0,0447 |
| Minimum | 0,0000 | 15,0000 | 654,0000 | 0,0000 |
| 25% | 3,0000 | 22,0000 | 2401,7800 | 0,0000 |
| 50% | 10,0000 | 26,0000 | 3130,9200 | 0,0000 |
| 75% | 30,0000 | 30,0000 | 3680,3500 | 0,0400 |
| Maksimum | 263,0000 | 50,0000 | 4939,9000 | 0,5385 |

Son olarak Şekil 4.9'deki grafikte de kullanıcı dağılımları toplu olarak görülmektedir. 3. kümeye yerleşen kullanıcıların yoğunluğu dikkat çekmektedir. Diğer kümelerdeki kullanıcı sayıları grafikte dikkat çekmeyecek düzeyde düşük çıkmıştır.



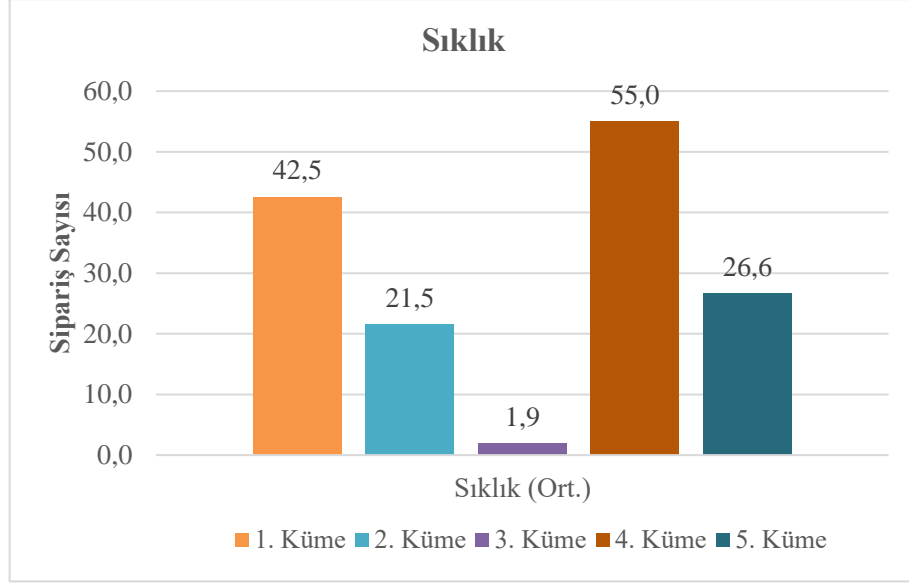
Şekil 4.9. Birch Hiyerarşik Kümeleme, Kullanıcı Sayılarının Kümelere Göre Dağılımları

Kriterlerin 5 küme için toplu grafikleri de aşağıdaki gibidir.



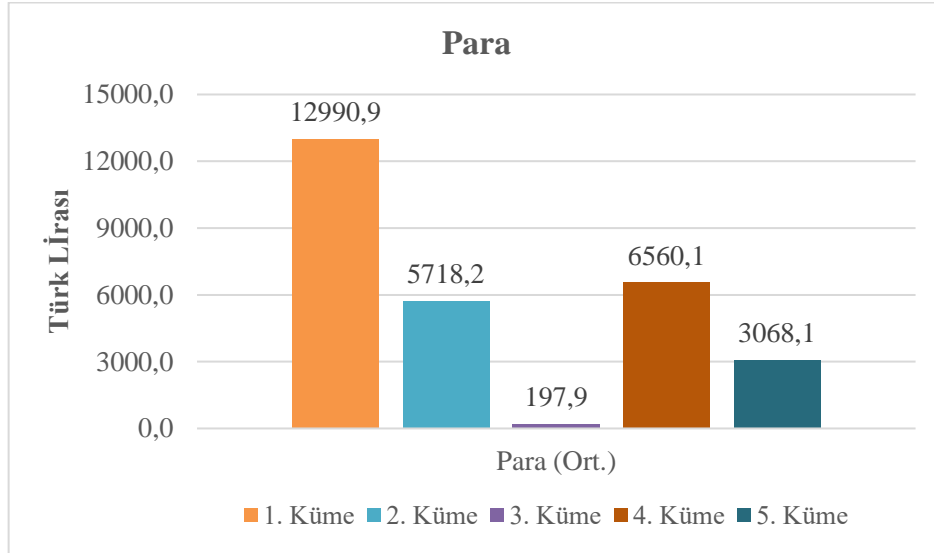
Şekil 4.10. Birch Hiyerarşik Kümeleme, Kümelere Göre Ortalama Güncellik Değerleri

Şekil 4.10'a göre, en güncel kullanıcılar 4. kümedeyken güncelliği en düşük olan kullanıcılar 3. kümede yer almışlardır.



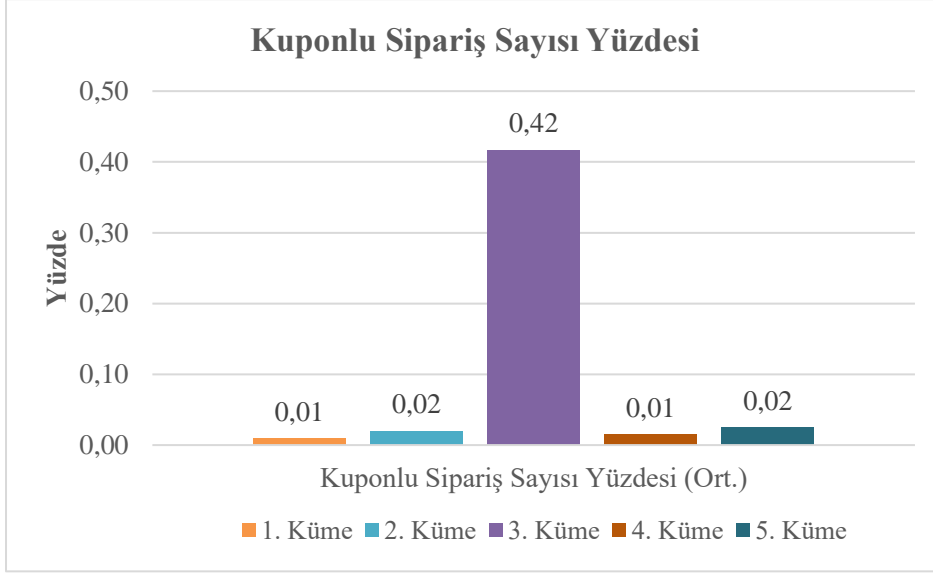
Şekil 4.11. Birch Hiyerarşik Kümeleme, Kümelere Göre Ortalama Sıklık Değerleri

Şekil 4.11'e göre, en sık sipariş veren kullanıcılar 4. kümede yer alırken, en az sipariş veren kullanıcılar 3. kümede yer almışlardır.



Şekil 4.12. Birch Hiyerarşik Kümeleme, Kümelere Göre Ortalama Para Değerleri

Şekil 4.12'ye göre, en yüksek sepet tutarlı kullanıcılar 1. kümede yer alırken en düşük sepet tutarlı kullanıcılar 3. kümede yer almışlardır.



Şekil 4.13. Birch Hiyerarşik Kümeleme, Kümelere Göre Ortalama Kuponlu Sipariş Sayısı Yüzdesi Değerleri

Şekil 4.13'e göre, kupon kullanım yüzdesi en yüksek olan kullanıcılar 3. kümede yer alırken, en düşük olan kullanıcılar 1. ve 4. kümelere yer almışlardır.

4.5.3. Gauss Karma Model Kümeleme

Gauss karma model kümeleme uygulamasında Kmeans, K-Means++, Random ve Random_From_Data parametreleri ve Full, Tied, Diag ve Spherical kovaryans tipleri kullanılarak kümeleme denemeleri yapılmıştır. Aşağıda bu 4 parametreye ve 4 kovaryans tipine göre sonuçların kıyaslandığı tablo yer almaktadır.

Çizelge 4.21'deki sonuçlar Calinski-Harabasz indeksi, Davies-Bouldin indeksi ve süre kriterlerine göre incelendiğinde Calinski ve Davies metodlarına göre parametrenin k-means, kovaryans tipinin tied olduğu versiyonun sonuçları daha iyi çıkmıştır. Model çözüm süresine bakıldığında ise parametrenin random, kovaryans tipinin ties olduğu model daha iyi çıkmıştır. Fakat süre kriteri arasında iki model arasında dikkate değer bir fark oluşmadığı için parametrenin k-means, kovaryans tipinin tied olduğu versiyonun kullanılması uygun bulunmuştur.

Çizelge 4.21. Gauss Karma Model Kümeleme Algoritma Kıyaslaması

| <i>Küme Sayısı</i> | <i>Parametreler</i> | <i>Kovaryans Tipi</i> | <i>Calinski-Harabasz Skoru</i> | <i>Davies-Bouldin Skoru</i> | <i>Süre (sn)</i> |
|--------------------|---------------------|-----------------------|--------------------------------|-----------------------------|------------------|
| 5 | kmeans | full | 54535,6588 | 2,2192 | 16,3163 |
| 5 | kmeans | tied | 158937,8832 | 0,8229 | 2,8346 |
| 5 | kmeans | diag | 104581,7303 | 1,2047 | 5,4891 |
| 5 | kmeans | spherical | 127015,2291 | 1,0082 | 6,7252 |
| 5 | k-means++ | full | 72903,6047 | 1,7736 | 13,5904 |
| 5 | k-means++ | tied | 84739,6583 | 1,2484 | 1,6287 |
| 5 | k-means++ | diag | 92413,0902 | 1,5284 | 4,9523 |
| 5 | k-means++ | spherical | 118327,2475 | 0,9366 | 5,3713 |
| 5 | random | full | 73624,1267 | 2,0601 | 18,3347 |
| 5 | random | tied | 22163,2463 | 1,7513 | 1,2021 |
| 5 | random | diag | 32134,9796 | 3,7505 | 6,0961 |
| 5 | random | spherical | 112946,2709 | 0,9686 | 11,1210 |
| 5 | random_from_data | full | 79142,4115 | 1,3730 | 11,6285 |
| 5 | random_from_data | tied | 73944,7669 | 1,3001 | 1,5437 |
| 5 | random_from_data | diag | 35415,0985 | 3,0628 | 7,8927 |
| 5 | random_from_data | spherical | 112544,9236 | 0,9705 | 5,7182 |

“K-means” parametresi ve “Tied” kovaryans tipiyle yapılan Gauss Karma Model kümelemenin küme özellikleri aşağıda verilmiştir.

- Çizelge 4.22 incelendiğinde, 1. kümeye 79.595 kullanıcının yerleştiği görülmektedir. Bu kullanıcıların güncellikleri ortalama 202 gün, ortalama sipariş verme sayıları 2 ve toplam sepet tutarları 137 TL çıkmıştır. Kuponlu sipariş verme yüzdeleri ise ortalama %0,4’dür. Bu kullanıcıların güncellik durumları ortalamada çok düşüktür, düşük sepet tutarlı kuponsuz siparişler verme eğiliminde oldukları tespit edilmiştir. Kuponsuz sipariş vermelerine rağmen sipariş vermeye devam etmemiş ve güncel olmayan kullanıcılarıdır.

Çizelge 4.22. Gauss Karma Model Kümeleme, 1. Kümenin Özellikleri

| <i>1. kümenin özellikleri</i> | <i>Güncellik (R)</i> | <i>Sıklık (F)</i> | <i>Para (M)</i> | <i>Kuponlu Sipariş Sayısı Yüzdesi</i> |
|-------------------------------|----------------------|-------------------|-----------------|---------------------------------------|
| Adet | 79595 | 79595 | 79595 | 79595 |
| Ortalama | 202,4130 | 1,5924 | 137,9531 | 0,0442 |
| Standart Sapma | 48,1316 | 1,2581 | 154,5735 | 0,1336 |
| Minimum | 120,0000 | 1,0000 | 20,0000 | 0,0000 |
| 25% | 160,0000 | 1,0000 | 51,9000 | 0,0000 |
| 50% | 198,0000 | 1,0000 | 85,7500 | 0,0000 |
| 75% | 239,0000 | 2,0000 | 157,7500 | 0,0000 |
| Maksimum | 299,0000 | 17,0000 | 2430,7300 | 0,5455 |

- Çizelge 4.23 incelendiğinde, 2. kümeye 57.853 kullanıcı yerleştiği görülmektedir. Bu kullanıcıların güncellikleri ortalama 64 gün, ortalama sipariş verme sayıları 1 ve toplam sepet tutarları 164 TL çıkmıştır. Kuponlu sipariş verme yüzdeleri ise ortalama %98'dir. Bu kullanıcıların güncellik durumları ortalamada düşük ve düşük sepet tutarlı kuponlu sipariş verme eğiliminde oldukları tespit edilmiştir. İlk kümedeki kullanıcılara göre daha günceldirler ve yine ilk kümedeki kullanıcıların tersine kuponlu sipariş verme oranları çok yüksektir. Bu kullanıcıların düşük sepet tutarlı, kuponlu ve bir defa sipariş verme durumları vardır.

Çizelge 4.23. Gauss Karma Model Kümeleme, 2. Kümenin Özellikleri

| <i>2. kümenin özellikleri</i> | <i>Güncellik (R)</i> | <i>Sıklık (F)</i> | <i>Para (M)</i> | <i>Kuponlu Sipariş Sayısı Yüzdesi</i> |
|-------------------------------|----------------------|-------------------|-----------------|---------------------------------------|
| Adet | 57853 | 57853 | 57853 | 57853 |
| Ortalama | 64,4691 | 1,3291 | 163,6493 | 0,9846 |
| Standart Sapma | 41,3390 | 0,8411 | 95,5032 | 0,0685 |
| Minimum | 0,0000 | 1,0000 | 25,0000 | 0,5714 |
| 25% | 23,0000 | 1,0000 | 107,9800 | 1,0000 |
| 50% | 68,0000 | 1,0000 | 130,2500 | 1,0000 |
| 75% | 96,0000 | 1,0000 | 174,2500 | 1,0000 |
| Maksimum | 155,0000 | 17,0000 | 1668,4500 | 1,0000 |

- Çizelge 4.24 incelendiğinde, 3. kümeye 108.168 kullanıcı yerleştiği görülmektedir. Bu kullanıcıların güncellikleri ortalama 56 gün, ortalama sipariş verme sayıları 2 ve

toplam sepet tutarları 257 TL çıkmıştır. Kuponlu sipariş verme yüzdeleri ise ortalama %0,7'dir. Bu kullanıcıların güncellik durumları diğer iki kümeye göre daha iyi durumda olmasına rağmen ortalama sayılabilir. Ortalama toplam sepet tutarı düşük olmasına rağmen kuponsuz sipariş verme eğiliminde oldukları tespit edilmiştir.

Çizelge 4.24. Gauss Karma Model Kümeleme, 3. Kümenin Özellikleri

| <i>3. kümenin özellikleri</i> | <i>Güncellik (R)</i> | <i>Sıklık (F)</i> | <i>Para (M)</i> | <i>Kuponlu Sipariş Sayısı Yüzdesi</i> |
|-------------------------------|----------------------|-------------------|-----------------|---------------------------------------|
| Adet | 108168 | 108168 | 108168 | 108168 |
| Ortalama | 56,6230 | 2,3376 | 257,0601 | 0,0739 |
| Standart Sapma | 40,3456 | 2,0519 | 274,7733 | 0,1632 |
| Minimum | 0,0000 | 1,0000 | 20,0000 | 0,0000 |
| 25% | 19,0000 | 1,0000 | 81,7500 | 0,0000 |
| 50% | 51,0000 | 2,0000 | 165,4500 | 0,0000 |
| 75% | 90,0000 | 3,0000 | 314,3000 | 0,0000 |
| Maksimum | 154,0000 | 19,0000 | 2563,0000 | 0,5833 |

- Çizelge 4.25 incelendiğinde, 4. kümeye 61.438 kullanıcı yerleştiği görülmektedir. Bu kullanıcıların güncellikleri ortalama 247 gün, ortalama sipariş verme sayıları 1 ve toplam sepet tutarları 139 TL çıkmıştır. Kuponlu sipariş verme yüzdeleri ise ortalama %98'dir. Bu kullanıcıların güncellik durumları diğer kümelere göre ortalamada düşük ve düşük sepet tutarlı kuponlu sipariş verme eğiliminde oldukları tespit edilmiştir. Bu kullanıcılar düşük sepet tutarlı, kuponlu ve bir defa sipariş vermektedirler.

Çizelge 4.25. Gauss Karma Model Kümeleme, 4. Kümenin Özellikleri

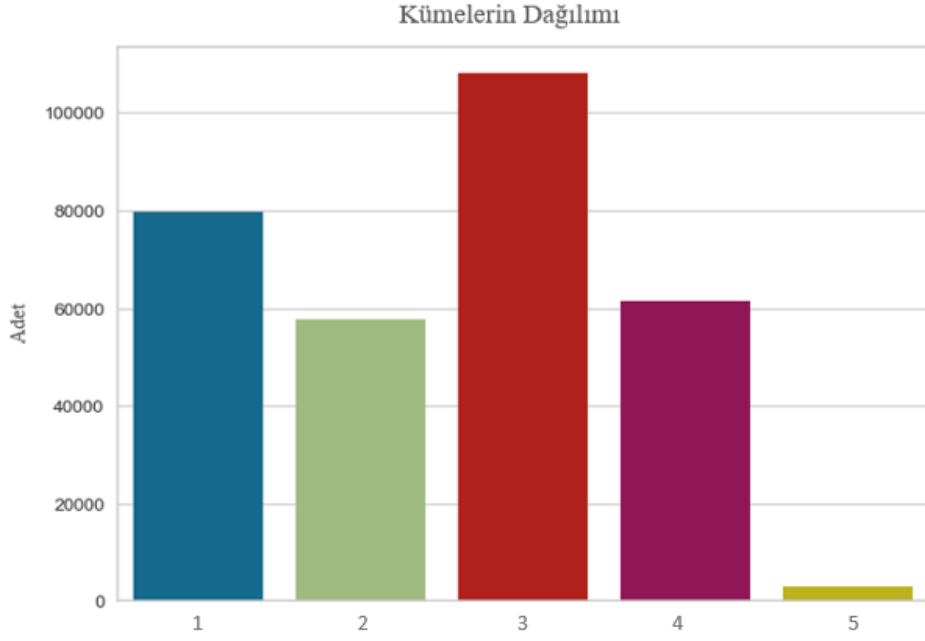
| <i>4. kümenin özellikleri</i> | <i>Güncellik (R)</i> | <i>Sıklık (F)</i> | <i>Para (M)</i> | <i>Kuponlu Sipariş Sayısı Yüzdesi</i> |
|-------------------------------|----------------------|-------------------|-----------------|---------------------------------------|
| Adet | 61438 | 61438 | 61438 | 61438 |
| Ortalama | 247,0770 | 1,4100 | 139,2149 | 0,9816 |
| Standart Sapma | 30,6807 | 0,8907 | 82,5763 | 0,0804 |
| Minimum | 116,0000 | 1,0000 | 20,2500 | 0,5000 |
| 25% | 224,0000 | 1,0000 | 93,9000 | 1,0000 |
| 50% | 241,0000 | 1,0000 | 105,2500 | 1,0000 |
| 75% | 279,0000 | 1,0000 | 150,2500 | 1,0000 |
| Maksimum | 299,0000 | 15,0000 | 1496,7500 | 1,0000 |

- Çizelge 4.26 incelendiğinde, 5. kümeye 3.042 kullanıcı yerleştiği görülmektedir. Bu kullanıcıların güncellikleri ortalama 37 gün, ortalama sipariş verme sayıları 18 ve toplam sepet tutarları 2831 TL çıkmıştır. Kuponlu sipariş verme yüzdeleri ise ortalama %0,3'dür. Bu kümede yer alan kullanıcıların diğer 4 kümeye göre güncel olduğu, sık sipariş verdikleri ve verdikleri siparişleri kuponsuz ve yüksek sepet tutarlı oluşturma eğiliminde oldukları söylenebilir.

Çizelge 4.26. Gauss Karma Model Kümeleme, 5. Kümenin Özellikleri

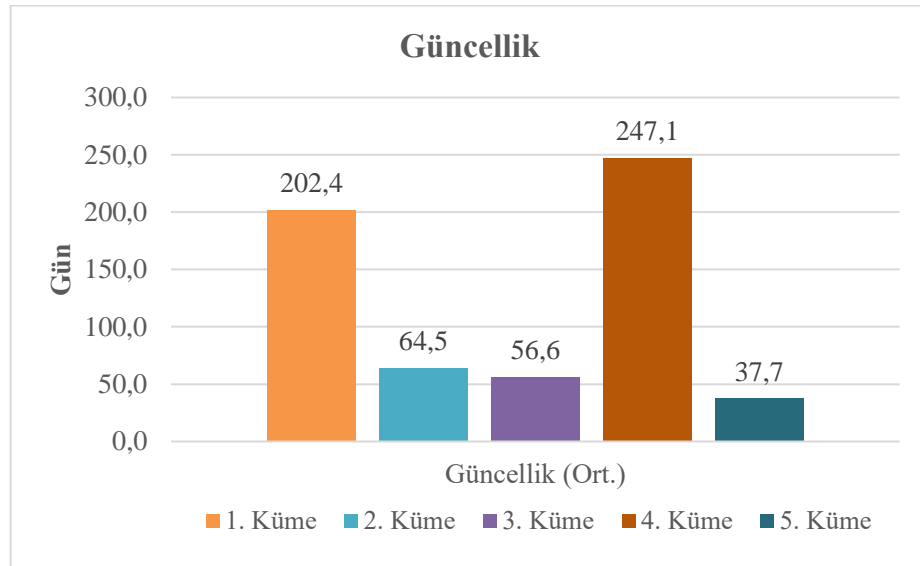
| <i>5. kümenin özellikleri</i> | <i>Güncellik (R)</i> | <i>Sıklık (F)</i> | <i>Para (M)</i> | <i>Kuponlu Sipariş Sayısı Yüzdesi</i> |
|-------------------------------|----------------------|-------------------|-----------------|---------------------------------------|
| Adet | 3042 | 3042 | 3042 | 3042 |
| Ortalama | 37,7242 | 18,2778 | 2831,1277 | 0,0338 |
| Standart Sapma | 46,9655 | 10,4037 | 1946,1250 | 0,0659 |
| Minimum | 0,0000 | 2,0000 | 487,0000 | 0,0000 |
| 25% | 5,0000 | 12,0000 | 1735,4425 | 0,0000 |
| 50% | 17,0000 | 16,0000 | 2284,1650 | 0,0000 |
| 75% | 51,0000 | 21,0000 | 3250,4125 | 0,0526 |
| Maksimum | 288,0000 | 112,0000 | 22172,1600 | 0,8500 |

Son olarak Şekil 4.14'deki grafikte de kullanıcı dağılımları toplu olarak görülmektedir. 5. küme haricinde diğer kümelerdeki kullanıcı sayılarının diğer metotlara göre daha dengeli dağıldığı söylenebilir. 3. küme kullanıcı sayısının en çok olduğu kümedir.



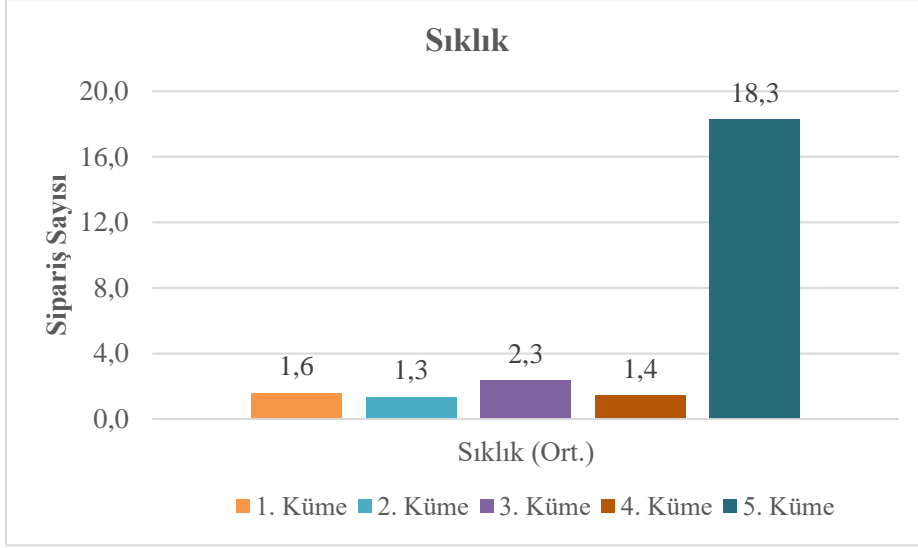
Şekil 4.14. Gauss Karma Model Kümeleme, Kullanıcı Sayılarının Kümelere Göre Dağılımları

Kriterlerin 5 küme için toplu grafikleri de aşağıdaki gibidir.



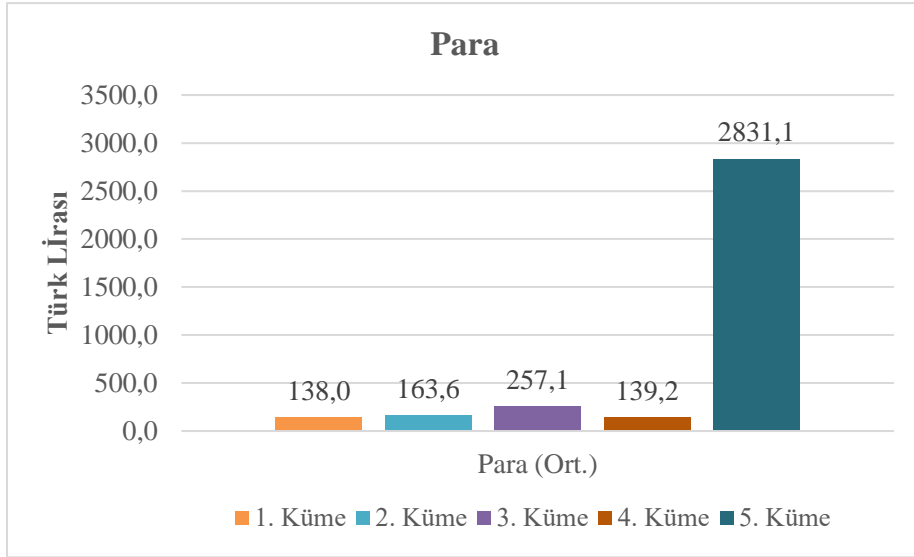
Şekil 4.15. Gauss Karma Model Kümeleme, Kümelere Göre Ortalama Güncellik Değerleri

Şekil 4.15'e göre, en güncel kullanıcılar 5. kümedeyken güncelliği en düşük olan kullanıcılar 4. kümede yer almıştır.



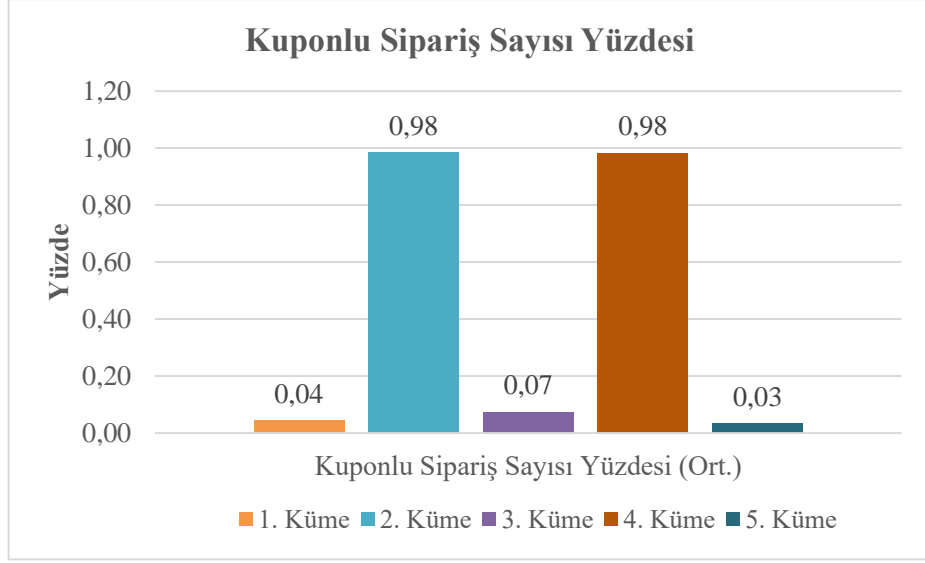
Şekil 4.16. Gauss Karma Model Kümeleme, Kümelere Göre Ortalama Sıklık Değerleri

Şekil 4.16'ya göre, en sık sipariş veren kullanıcılar 5. kümede yer alırken, en az sipariş veren kullanıcılar 2. kümede yer almışlardır.



Şekil 4.17. Gauss Karma Model Kümeleme, Kümelere Göre Ortalama Para Değerleri

Şekil 4.17'ye göre, en yüksek sepet tutarlı kullanıcılar 5. kümede yer alırken en düşük sepet tutarlı kullanıcılar 1. kümede yer almışlardır.



Şekil 4.18. Gauss Karma Model Kümeleme, Kümelere Göre Ortalama Kuponlu Sipariş Sayısı Yüzdesi Değerleri

Şekil 4.18'e göre, kupon kullanım yüzdesi en yüksek olan kullanıcılar 2. ve 4. kümelerde yer alırken, en düşük olan kullanıcılar 5. kümede yer almışlardır.

4.6. Metotların Sonuçlarının Karşılaştırması

Kümeleme algoritmalarından K-Ortalamalar, Birch Hiyerarşik ve Gauss Karma Modeli denenmiştir. Bu modellerin üstteki bölümde kendi içlerinde parametre kıyaslamaları yapılmıştır. En iyi sonuç veren parametrelili kümelemelerin sonuçları da denenilen diğer kümeleme yöntemlerinin sonuçlarıyla kıyaslanmıştır. Kıyaslama için Calinski-Harabasz İndeksi, Davies-Bouldin İndeksi ve modelin çalışma süresi kriterleri değerlendirilecektir.

Çizelge 4.27. Kümeleme Modellerinin Performans Kriterlerine Göre Kıyaslanması

| <i>Kümeleme Modelleri</i> | <i>Calinski-Harabasz Skoru</i> | <i>Davies-Bouldin Skoru</i> | <i>Süre (sn)</i> |
|---------------------------------|--------------------------------|-----------------------------|------------------|
| K-Ortalamlar (Elkan) | 166234,8589 | 0,7743 | 4,8087 |
| Birch Hiyerarşik Kümeleme | 22309,1025 | 0,8882 | 26,6687 |
| Gauss Karma Model (Kmeans,Tied) | 158937,8832 | 0,8229 | 2,8346 |

Çizelge 4.27’de yer alan sonuçlara göre, Calinski skoru ve Davies skoru en iyi sonuçları K-Ortalamlar kümeleme sonuçlarında vermiştir. Bu iki değerlendirme kriterine göre ikinci sırada Gauss Karma Model kümeleme yönteminin sonuçları yer almıştır. Birch Hiyerarşik Kümenin bu iki değerlendirme kriterine göre sonuçları en kötü olarak çıkmıştır.

Süre kriterine bakıldığında ise en iyi sonucu Gauss Karma modeli vermiştir. Onu süre olarak sırasıyla K-Ortalamlar kümeleme ve Birch Hiyerarşik Kümeleme takip etmiştir. Yine de Gauss Karma ile K-Ortalamlar’ın sürelerinde dikkate değer bir fark gözlemlenmemiştir.

Birch hiyerarşik kümeleme üç performans kriterine göre de son sırada yer almıştır. Yukarıda küme tanımlamalarını paylaştığımız bölümde de bu metodun sağlıklı kümelemeler yapamadığı görülmüştür. K-Ortalamlar ve Gauss Karma modelleri birbirine yakın sonuçlar vermiştir. Bu noktada hangi modelin sonuçlarının kullanılması gerektiğine pratik karşılaştırmalar yaparak yetkililer ile kararlaştırılmasına karar verilmiştir.

Kümeleme modellerine göre elde edilen sonuçlar şirket pazarlama yetkililerinin görüşlerine sunulmuştur. Sonuçlardan Birch Hiyerarşik Kümeleme modelinin verdiği sonuçlar uygun bulunmamıştır. K-Ortalamlar ve Gauss Karma Model sonuçları aşağıda tek tabloda toplanmış ve incelenmiştir.

Çizelge 4.28. K-Ortalamlar ve Gauss Karma Model Kümeleme Sonuçlarının
Kümelerin Özelliklerine Göre Kıyaslanması

| <i>K-Ortalamlar Kümeleme Sonuçları</i> | | | | | <i>Gauss Karma Model Kümeleme Sonuçları</i> | | | |
|--|------------------|---------------|-------------|--------------------------------------|---|---------------|-------------|--------------------------------------|
| | <i>Güncellik</i> | <i>Sıklık</i> | <i>Para</i> | <i>Kupon Kullanım Sayısı Yüzdesi</i> | <i>Güncellik</i> | <i>Sıklık</i> | <i>Para</i> | <i>Kupon Kullanım Sayısı Yüzdesi</i> |
| 1. Küme | 53 | 7 | 843 | 10% | 202 | 2 | 137 | 0,4% |
| 2. Küme | 127 | 2 | 141 | 2% | 64 | 1 | 164 | 98% |
| 3. Küme | 29 | 24 | 3996 | 2% | 56 | 2 | 257 | 0,7% |
| 4. Küme | 62 | 1 | 172 | 92% | 247 | 1 | 139 | 98% |
| 5. Küme | 246 | 1 | 141 | 95% | 37 | 18 | 2831 | 3% |

Çizelge 4.28 yetkililerle incelenmiş ve özellikle sıklık, para ve kupon kullanım sayısı yüzdesi özelliklerine göre K-Ortalamlar tekniğiyle kümelemenin strateji çalışmalarına uygun olacak şekilde sonuçlar verdiğine karar verilerek K-Ortalamlar kümeleme metodunun sonuçlarının kullanım için seçilmesine karar verilmiştir.

4.7. Segmentlere Göre Müşteri İlişkileri Geliştirme Önerileri

Pazarlama yönetimi, örgütsel başarının sağlanması amacıyla hedef pazarlarla yararlı ticaret yapmak ve sürdürmek için planlanan ürün ve hizmet modellerinin tasarlanması, tahmin edilmesi ve sürdürülmesi dahil olmak üzere inceleme, düzenleme, yürütme, planlama ve kontrol etme yöntemidir. Pazarlama segmentasyonu kavramı, ürün veya hizmet talep eden müşterilerin ayırt edici özelliklerinin olduğunun kabul edilmesinden kaynaklanmaktadır.

Firmalar kaynaklarını doğru yönlendirebilmek ve özellikle finansal konularda verimlilik sağlayabilmek için müşterilerini segmentlere ayırmaya ihtiyaç duyarlar. Bu çalışmada temelde amaçlanan elde edilen sonuçlarla daha etkin müşteri segmentleri oluşturarak pazarlama bölümüne bilgi sağlamaktır. Çünkü müşterilerin potansiyelini anlamak ve en yüksek fırsatın olduğu uzun vadeli ilişkiler geliştirmeye para yatırmak çok önemlidir.

Yapılan segmentasyon çalışmasına göre bir üst bölümde K-Ortalamalar kümeleme metodunun sonuçları kullanılmaya karar verilmiştir. Bu metodun özniteliklerine göre tekrar bir inceleme yapıldığında şu özet bilgiler ortaya çıkmaktadır:

- Güncellik: En güncel müşterilerin olduğu küme ortalama 29 gün ile 3. kümedir. Bunu sırasıyla ortalama 53 gün ile 1. küme, ortalama 62 gün ile 4. küme, ortalama 127 gün ile 2. küme ve ortalama 246 gün ile 5. küme takip etmektedir.
- Sıklık: Sipariş sıklığı en yüksek olan küme ortalama 24 toplam sipariş ile 3. kümedir. Bunu sırasıyla ortalama 7 toplam sipariş ile 1. küme, ortalama 2 toplam sipariş 2. küme ve ortalama 1 toplam sipariş ile 4 ve 5. kümeler takip etmektedir.
- Para: Toplam sepet tutarına göre en yüksek olan küme ortalama 3996 TL ile 3. kümedir. Bunu sırasıyla ortalama 843 TL ile 1. küme, ortalama 172 TL ile 4. küme ve ortalama 141 TL ile 2 ve 5. kümeler takip etmektedir.
- Kupon Kullanım Sayısı Yüzdesi: Kupon kullanım sayısı yüzdesine göre en az kupon kullanımıyla sipariş veren küme ortalama %2 ile 2. ve 3. kümelerdir. Bunu sırasıyla ortalama %10 ile 1. küme, ortalama %92 ile 4. küme ve ortalama %95 ile 5. küme takip etmektedir.

Kümelerin bu özelliklerine göre segmentlerine ayırıp müşteri ilişkileri geliştirme stratejileri önerilmiştir.

- 1. Segment: 4 öznitelik için de en iyi kullanıcıların olduğu 3. Küme, 1. segment olarak belirlenmiştir. Bu segmentteki kullanıcılar şirkete sadık olup kupon beklentisi olmadan sık ve yüksek sepet tutarlı siparişler vermektedirler. Bu kullanıcılar için ekstra kupon çalışması yapılmasına gerek yoktur fakat bağlılık düzeylerinin düşmesine sebep olacak operasyonel hataların yapılmaması ve memnuniyet düzeylerinin azalmadan devam etmesi sağlanmalıdır. Bunu sağlayabilmek için önceki siparişlerinin sepet analizleri yapılarak bu ürünlerin özellikle kampanyalı olduğu dönemlerde kullanıcıya bildirimler gönderilmesi sağlanabilir. Bu kullanıcılara belirli

aralıklarla birebir telefon kanalı aracılığıyla anketler yapılarak talepleri toplanabilir ve bu taleplere göre aksiyonlar alınması sağlanabilir, çünkü bu tip kullanıcıların belirli bir bağlılık düzeyleri olduğu için birebir iletişim kurma ve taleplerinin bu şekilde dinlenmesinin kendilerini değerli hissetmeleri noktasında önemli olduğuna dair çalışmalar mevcuttur. Siparişleriyle alakalı soru, talep ve şikayetleri için şirketin müşteri hizmetleri bölümüne bağlanmaları durumunda ilk öncelik verilebilir ve karşılama için 10 saniye veya daha az bir süre hedefi konabilir. Özellikle telefon harici kanallardan (chat, sosyal medya, şikâyet platformları vs.) iletilen şikayetlerde kesinlikle telefon kanalından ulaşılarak şikayetleriyle alakalı telafilerin yapılması sağlanabilir. Bu kullanıcılara siparişlerinin tesliminde gecikme veya iptal yaşanmaması adına siparişi kuryeye atama noktasında öncelik verilebilir ve teslimat ücreti almamak gibi memnuniyet artırıcı aksiyonlar alınabilir. Çünkü bu sektöre sıklıkla dile getirilen mükemmel sipariş yani eksiksiz/sorunsuz ve tam zamanında giden siparişlerin bu kullanıcı segmenti için önemi çok fazladır. Siparişleriyle alakalı gecikme veya iptal durumları yaşanırsa şikâyet iletmelerine fırsat kalmadan düzenli takip yaparak taraflarına telafi kuponları sunulabilir. Yaşadıkları olumsuz deneyim neticesinde şikâyet dahi etmeden şirketi tercih etmeyi bırakma riskleri olduğu için düzenli takip bu segmentin siparişleri için çok önemlidir. Siparişlerindeki ürünlerle alakalı hasar bildirimini yapmaları durumunda sorgusuz ürün tedariki ve yaşanan durumun telafisi için kupon tanımlamaları yapılabilir. Lojistik maliyetleri dikkate alındığında genelde ürünün tekrar tedariki tercih edilen bir aksiyon olmamakla birlikte bu kullanıcı segmenti için kupon tanımlamalarının yanında sorun yaşadıkları ürünün yeniden tedarikinin sağlanması kullanıcıların olumsuz deneyimini ortadan kaldıracak bir hamle olacaktır. Şirketlerin yaptığı sosyal sorumluluk projelerine davet edilebilir veya isimlerinin anılması sağlanabilir. (Adına ağaç dikimi vs. gibi)

- 2. Segment: Sıklık, güncellik ve para özniteliklerinde 3. kümeden sonra en iyi kullanıcıların olduğu 1. küme, 2. segment olarak belirlenmiştir. 2. segment kullanıcıları için potansiyel 1. segment müşterileridir denebilir, bu yüzden önemli bir segmenttir. Bu segmentteki kullanıcıların bir kısmı kupon kullanımına önem verdiği için bu segment özelinde sipariş sıklığını da artırmak amacıyla özel kupon çalışmaları yapılabilir. Yine bu segmentte yer alan kullanıcıların soru, talep ve şikayetlerini iletmek için müşteri hizmetlerine ikinci öncelikle bağlanmaları sağlanabilir.

Siparişleriyle alakalı şikâyet veya ürün hasar bildirimini ilettiklerinde ivedilikle telafisi sağlanabilir. Siparişleriyle alakalı gecikme veya iptal durumları yaşanırca şikâyet iletmelerine fırsat bırakmadan düzenli takip yaparak taraflarına telafi kuponları sunulabilir. Sipariş sıklıklarını artırabilmek için birbirine bağlı kupon çalışmaları yapılabilir. Bu tarz çalışmalar sadakat programları kapsamında anılır ve kullanıcıların sipariş tekrarını sağlayacak nitelikte yapılan çalışmaların tümünü kapsamaktadır. Bu kullanıcıların sipariş özelinde doldurdukları anketler düzenli takip edilebilir ve sorun tespit edildiğinde kupon telafisi sağlanabilir.

- 3. Segment: Kupon kullanım sayısı yüzdesine göre 1. kümeden ve sıklık durumuna göre 4 ve 5. Kümelerden daha iyi kullanıcıların olduğu 2. Küme, 3. Segment olarak belirlenmiştir. Bu kullanıcıların kuponsuz sipariş verme eğilimleri vardır ancak güncellik durumları iyi değildir. Birden fazla ve kuponsuz sipariş verme eğiliminde oldukları için kaybedilmemelidirler ve sipariş vermelerini tetikleyecek çalışmalar yapılmalıdır. Bunun için kullanıcıya sık sık bildirim göndererek sipariş verme hatırlatması yapılabilir. Sepet tutarlarını artırabilmek için önceki siparişlerinden sepet analizleri yapılabilir ve verdikleri ürünleri tamamlayacak veya benzer olan ürünleri sepetlerine atmalarını sağlayacak ek ürün öneri sistemleri üzerine çalışılabilir. Kullanıcıların siparişleriyle alakalı verdiği puanlar, yorumlar veya doldurdukları anketler belirli aralıklarla takip edilebilir, telafi noktasında kupon verilmesi sağlanabilir. Şirkete bağlılık düzeyi hassas durumda olan bu kullanıcıların sadakat çalışmaları yapılarak birbirini takip eden siparişler vermesi sağlanabilir.
- 4. Segment: Güncellik ve para özniteliklerine göre 2. kümeden, kupon kullanım sayısı yüzdesi özniteliğine göre 5. kümeden daha iyi olan kullanıcıların yer aldığı 4. küme, 4. segment olarak belirlenmiştir. Bu kullanıcılar daha güncel fakat çoğunlukla tek sipariş vermiş kullanıcılardır. Bu kullanıcıların büyük çoğunluğu yeni kullanıcılardır ve şirkete sadakati sağlayabilmek için ilk siparişlerinde yaşadıkları deneyim ve takip edecek olan siparişler için yapılacak çalışmalar önemlidir. Bu kullanıcıların bir kısmı muhtemelen ilk sipariş kuponlarını kullanıp tekrar sipariş vermeyip 5. kümeye dahil olacaklardır, fakat devam siparişlerini veren kullanıcıların yüzdelerini artırmak için sadakat çalışmaları yapılarak sipariş vermeleri tetiklenebilir ve daha üst segment kullanıcılar olmaları sağlanabilir. İlk kez sipariş veren kullanıcıların siparişleri

operasyonel olarak önem arz ettiđi için yine ilk kümedeki kullanıcılara önerildiđi gibi sipariřlerinin kurye atamaları önceliklendirilebilir. İlk sipariřlerinde yařadıkları deneyimin nasıl olduđunu anlamak için daha özel anket çalıřmaları yapılabilir ve kötü deneyim yařayan müşteriler tespit edilerek telafi yapılması noktasında birebir telefon kanalıyla temas kurularak yařadıkları kötü deneyim için özür dilenip kupon sunulabilir. Bu kullanıcıların devam sipariřlerini vermesi için belirli aralıklarla bildirim gönderimi yapılarak sipariř verme hatırlatması yapılabilir.

- 5. Segment: 4 öznitelik için de en kötü olan kullanıcıların yer aldığı 5. Küme, 5. segment olarak seçilmiştir. Bu kullanıcılar güncel değildir ve kupon kullanım yüzdelerinin %95 olması göz önüne alındığında ilk sipariř kuponunu kullanmak amacıyla sipariř verdikleri söylenebilir. Bu kullanıcılara kayıp müşteriler de denebilir. Büyük tutarlı kupon olması durumunda sipariř verme olasılığı olan bu kullanıcılara bu tarz bir kupon çalıřması yapıldığı takdirde bildirimler gönderilebilir ve tekrar sipariř vermeleri sağlanabilir. Fakat kaynakların sınırlı olduđu günümüzde bu kullanıcı grubu için kaynak ayırmak riskli olabilir. Bu kullanıcılar kaynak kullanımını noktasında göz ardı edilebilirler.

5. SONUÇ VE TARTIŞMA

Veri madenciliğinin pazarlama alanında kullanımı çok yaygındır. Şirketler, veri madenciliğini pazarlama kapsamında müşteri ilişkileri yönetimi, müşteri edinme, müşteri segmentasyonu, kampanya yönetimi, mevcut müşterileri elde tutma, sepet analizleri gibi pek çok alanda kullanmaktadır. Benzer davranışları olan müşterilerin gruplandırılması, şirketlere hedef pazarı daha iyi anlama fırsatı verdiği için, müşterilerin yönetilebilir bölümlere ayrılması şirketlerin varlığını sürdürebilmesi için hayati önem taşımaktadır.

Bu kapsamda bu çalışmada FMCG sektöründe yer alan bir online satış platformunun verilerinden faydalanılarak bu tip verilerde literatürde yer alan RFM kriterlerinin yanı sıra hangi öznitelik kullanmaya değer olduğu tespit edilmek istenmiştir. Öznitelik seçimi için Calinski-Harabasz İndeksi ve Davies-Bouldin İndeksi kullanılmıştır ve uygulama bölümünde yapılan çalışmalarla RFM kriterlerine ek olarak Kuponlu Sipariş Sayısı Yüzdesi kriterinin kullanılması gerektiği bulunmuştur.

Öznitelikler belirlendikten sonra belirlenen dört öznitelik kullanılarak kaç adet küme ile çalışılması gerektiği hesaplanmıştır. Küme sayısının tespitinde Elbow Yöntemi kullanılmıştır. Daha sonra şirket yetkililerinden alınan görüşle de çıkan sonuç pekiştirilerek küme sayısı beş olarak belirlenmiştir.

Bulunan öznitelikler ve küme sayısı ile K-Ortalamalar, Birch Hiyerarşik Kümeleme ve Gauss Karma Model metotları kullanılarak kümeleme yapılmıştır. Elde edilen kümelerin özellikleri öznitelikler bazında değerlendirilmiştir. Kullanılan kümeleme metotların sonuçları üç kriter ile kıyaslanmıştır. Bunlar Calinski-Harabasz İndeksi, Davies-Bouldin İndeksi ve modelin çözüm süresidir. Calinski-Harabasz ve Davies-Bouldin skorlarına göre K-Ortalamalarla yapılan kümeleme çalışması en iyi sonucu vermiştir. Modelin çözüm süresine bakıldığında ise en iyi sonucu Gauss Karma Model vermiştir.

Birch Hiyerarşik Kümeleme'nin sonuçları diğer iki kümeye göre çok farklı çıkmış ve başarılı bulunmamıştır. Bu şekilde bir sonuç çıkmasının sebebinin algoritmanın adım adım birleştirme işlemi yapması ve birleştirdiği kümeler üzerinde tekrar işlem yapmaması

olabileceği düşünülmektedir. Ayrıca Bocci ve Mingo [69], yaptıkları çalışmada Birch algoritmasında yaşanabilecek bir kirlenme durumundan bahsetmişlerdir ve bunu küme özelliklerinden farklı elemanların olmasıyla o kümenin özelliklerinin etkilenmesi olarak tanımlamışlardır. Bu durumdan, algoritma kullanımı için negatif bir özellik olarak bahsetmişlerdir.

K-Ortalamalar ve Gauss Karma Model'in performans kriterlerinin sonuçları dikkate alınarak şirket pazarlama yetkilileri ile görüşülmüştür ve K-Ortalamalar metodunun kümeleme sonuçlarının kullanılmasına karar verilmiştir.

Daha sonra bu metodun kümeleme sonuçlarına göre her segment için müşteri ilişkileri geliştirme önerilerinde bulunulmuştur. Bu segmentlerde 1. segment güncel, düzenli, yüksek sepet tutarlı ve kupon beklentisi olmadan sipariş veren kullanıcılardan oluşmaktadır. Bu kullanıcılar en önemli kullanıcılar olduğu için müşteri memnuniyetini üst düzeyde tutacak önerilerde bulunulmuştur. 2. segmentte yer alan kullanıcılar ise birinci segmente en yakın satın alma davranışları göstermektedir ve bu kullanıcıların da şirketi birinci segmentteki müşteriler kadar tercih etmesi için önerilerde bulunulmuştur. 3. Segmentte yer alan kullanıcılar şirketi birden fazla kez tercih etmiş fakat güncel olmayan kullanıcılardan oluşmaktadır. Bu segment kullanıcılarının kupon kullanabilmek için şirketi tercih etmediği göz önüne alındığında kaybedilmemesi gereken bir müşteri grubudur. 4. Segmentte yer alan kullanıcılar şirketi ilk kez tercih edilen kullanıcılardır ve yaşadıkları alışveriş deneyiminin analiz edilmesi ve yaşanmış olabilecek olumsuz bir deneyimin iyileştirilmesi büyük önem arz etmektedir. 5. ve son segmentte şirketi bir defa tercih etmiş, kupon kullanım oranı çok yüksek olan ve güncellikleri en kötü olan kullanıcılar yer almaktadır. Bu kullanıcıların ilk alışverişlerini ilk sipariş kuponu sebebiyle vermiş olmaları olasılığı yüksektir. Bu yüzden bu kullanıcılara fazladan pazarlama kaynağı ayırmaya gerek duyulmamıştır.

Son olarak, segmentasyon işleminin uygulaması yapıldıktan sonra yakından takip edilmesi ve belirli aralıklarla tekrarlanması gereken bir işlem olduğu söylenebilir. Her segment zamanla yeni gelen müşteri verileriyle evrileceği için bir defaya mahsus yapılması şirketlere başarı getirmemektedir. Ayrıca, bir segmente dahil edilen müşteri zaman içerisinde alışveriş davranışının değişmesiyle farklı segmentlere kayabilir. Bu

sebeplerle önceden belirlenen aralıklarla takip ve kontrol yapılması, segment özelliklerinin güncellenmesi ve tespit edilen deęişimlere göre önlem veya stratejik adımlar planlanması çok önemlidir.

Bunların yanı sıra, özellikle bu tez çalışmasında yapılan etiketlenmemiş verilerin denetimsiz makine öğrenmesi tekniklerinden kümeleme algoritmaları kullanılarak etiketlenmesi işleminden sonra, yeni gelen müşteri verilerinin denetimli makine öğrenmesi tekniklerinden sınıflandırma algoritmaları kullanılarak ilgili segmentlere yerleştirilmesi de uygun bir yaklaşım olabilir ve gelecek çalışmalarda ele alınabilir. Ayrıca, bütün bu işlemlerin kolaylıkla otomatik olarak yapılabileceęi, takip edilebileceęi ve pazarlama kaynaklarının etkin bir şekilde kullanılabilceęi karar destek sistemlerinin kurulması da şirketlere çok katkı sağlayacaktır.

6. KAYNAKLAR

- [1] F. K. Çağdaş, «Hızlı Tüketim Ürünleri (FMCG) Sektörü ve E-Ticaret,» 2020. [Çevrimiçi]. Available: <https://worldef.net/hizli-tuketim-urunleri-fmcg-sektoru-ve-e-ticaret/>. [Erişildi: 07. 11. 2022].
- [2] Y. Durmaz, Tüketici Davranışı, Ankara: Detay Yayıncılık, s.214, 2008, p. 214.
- [3] M. K. Tiwari, «An Empirical Analysis of Effect of Advertising on Marketing of FMCG Product,» *International Journal of Marketing and Technology*, pp. 167,200, 2(2012).
- [4] Anonim1, «Hızlı tüketim ürünleri sektörü yüzde 85 büyüdü,» 2022. [Çevrimiçi]. Available: <https://www.inbusiness.com.tr/sectorler/tarim-gida-perakende/2022/11/10/hizli-tuketim-urunleri-sektoru-yuzde-85-buyudu>. [Erişildi: 24 Aralık 2022].
- [5] B. Sohrabi ve A. Khanlari, «Customer Lifetime Value (CLV) Measurment based on RFM Model,» *Iranian Acc. Aud. Rev.*, no. 14, pp. 7-20, 2007.
- [6] A. Hughes, «Boosting Response with RFM,» *Marketing Tools*, pp. 4-10, 1996.
- [7] J. T. Wei, M. C. Lee, H. K. Chen ve H. H. Wu, «Customer Relationship Management in the Hairdressing Industry: An Application of Data Mining Techniques,» *Expert Systems with Applications*, 40, pp. 7513-7518, 2013.
- [8] F. Tao, «Customer Relationship Management Based on Increasing Customer Satisfaction,» *International Journal of Business and Social Science*, cilt 5, no. 5, 2014.
- [9] K. Hjort, B. Lantz, D. Ericsson ve J. Gattorna, «Customer Segmentation Based On Buying And Returning Behaviour,» *International Journal of Physical Distribution & Logistics Management, Vol. 43*, pp. 852-865, 2013.
- [10] G.-N. Vicente, G.-G. Hermenegildo ve O.-B. Raul , «Customer Relationship Management and Its Impact on Innovation: A literature Review,» *Journal of Business Research*, no. 129, pp. 83-87, 2021.

- [11] İ. D. Karakoç and P. Özkan, "Sağlık Sektöründe LRFM Analizi ile Pazar Bölümlendirme," in *PPAD Pazarlama Kongresi*, Kuşadası, Türkiye, 1-4 Mayıs 2019.
- [12] S. R. A. Ahmed, I. Al_Barazanchi ve Z. A. Jaaz, «Clustering Algorithms Subjected to K-mean and Gaussian Mixture Model on Multidimensional Data Set,» *Periodicals of Engineering and Natural Sciences*, cilt 7, pp. 448-457, 2019.
- [13] C. H. Cheng ve Y. S. Chen, «Classifying the Segmentation of Customer Value via RFM Model and RS Theory,» *Expert Systems with Applications*, 36, pp. 4176-4184, 2009.
- [14] M. Tavakoli, M. Mobini, M. Molavi and S. Etemad, "Customer Segmentation and Strategy Development based on User Behavior Analysis, RFM model and Data Mining Techniques: A Case Study," in *IEEE 15th International Conference on e-Business Engineering (ICEBE)*, 2018.
- [15] A. Çalış ve K. Baynal, «Kümeleme Analizi ile Bankacılık Sektöründe Satış Stratejilerinin Belirlenmesi,» *Beykent Üniversitesi Fen ve Mühendislik Bilimleri Dergisi*, 9(1), pp. 13-41, 2016.
- [16] J.-T. Wei, S.-Y. Lin ve H.-H. Wu, «A review of the application of RFM model,» *African Journal of Business Management*, cilt 4(19), pp. 4199-4206, 2010.
- [17] Z. Zalaghi ve Y. Varzi, «Measuring Customer Loyalty Using an Extended RFM and Clustering Technique,» *Management Science Letters*, pp. 905-912, 2014.
- [18] R. Soeini, "Customer Segmentation Based On Modified RFM Model in The Insurance Industry," in *Proceedings of 2012 4th International Conference on Machine Learning and Computing*, 2012.
- [19] J.-T. Wei, S.-Y. Lin, C.-C. Weng ve H.-H. Wu, «A Case Study of Applying LRFM Model in Market Segmentation of a Children's Dental Clinic,» *Expert Systems with Applications*, no. 39, p. 5529–5533, 2013.
- [20] J. N. Sari, L. E. Nugroho ve R. Ferdiana, «Review on Customer Segmentation Technique on Ecommerce,» *Advanced Science Letters*, cilt 4, no. 400–407, 2011.

- [21] M. Mihrimah Özmen, Y. Delice ve E. Kızılkaya Aydoğan, «Telekomünikasyon Sektöründe PSO ile Müşteri Bölümlemesi,» *Bilişim Teknolojileri Dergisi*, cilt 11, no. 2, 2018.
- [22] M. Namvar, M. R. Gholamian and S. Khakabi, "A Two Phase Clustering Method for Intelligent Customer Segmentation," in *International Conference on Intelligent Systems, Modelling and Simulation*, 2010.
- [23] C.-C. H. Chan, C.-B. Cheng ve W.-C. Hsien, «Pricing and Promotion Strategies of an Online Shop Based on Customer Segmentation and Multiple Objective Decision Making,» *Expert Systems with Applications* 38, pp. 14585-14591, 2011.
- [24] E. W. Lee, «Data Mining Application in Customer Relationship Management for Hospital Inpatients,» *Healthc Inform Res.*, cilt 18(3), pp. 178-185, 2012.
- [25] C. P. Ezenkwu, S. Ozuomba ve C. Kalu, «Application of K-Means Algorithm for Efficient Customer Segmentation: A Strategy for Targeted Customer Services,» *International Journal of Advanced Research in Artificial Intelligence*, Vol.4, p. 10, 2015.
- [26] M. Başkol, «RFM ve Uyum Analizi Kullanılarak Müşteri Segmentasyonunun Belirlenmesi,» *BMIJ*, pp. 8(4):902-938, 2020.
- [27] J.-T. Wei, S.-Y. Lin and H.-H. Wu, "A review of the application of RFM model," *African Journal of Business Management*, vol. 4(19), pp. 4199-4206, 2010.
- [28] A. Rachid, A. Amine, B. Bouikhalene ve R. Lbibb, «Customer Segmentation Model in E-commerce Using Clustering Techniques and LRFM Model: The Case of Online Stores in Morocco,» *International Scholarly and Scientific Research & Innovation*, cilt 9(8), 2015.
- [29] T. Yousefi, M. S. Odabaş ve R. Oktaş, «Kümeleme Algoritmalarında Kullanılan Farklı Yöntemlere Genel Bakış,» *Black Sea Journal of Engineering and Science*, cilt 3(4), pp. 173-189, 2020.
- [30] D. Meyer ve C. Cannon, *Building a Better Data Warehouse*, USA: Prentice Hall, 1998, pp. 20-21.

- [31] S. Ganesh, "Data Mining: Should It Be Included in The ‘Statistics’ Curriculum?," in *The Sixt International Conference on Teaching Statistics*, Güney Afrika, 2002.
- [32] M. Rahman, M. I. Ahmed and M. S. Hossain, "Analysis of Student’s Achievement through Educational Data Mining," in *International Conference on Information Systems and Advanced Technologies (ICISAT)*, Algeria, 2021.
- [33] P. Gray ve H. Watson, «Decision Support in the Data Warehouse,» *Prentice Hall PTR, New Jersey*, 1998.
- [34] U. T. Şimşek, *Veri Madenciliği ve Müşteri İlişkileri Yönetiminde (CRM) Bir Uygulama*, İstanbul: İstanbul Üniversitesi Sosyal Bilimler Enstitüsü, 2006.
- [35] G. Firican, «The 10 Vs of Big Data,» 2017. [Çevrimiçi]. Available: <https://tdwi.org/articles/2017/02/08/10-vs-of-big-data.aspx>. [Erişildi: 27 Aralık 2022].
- [36] S. Savaş, «Veri Madenciliğinin Tarihi ve Kullanıldığı Alanlar,» 18 Mayıs 2020. [Çevrimiçi]. Available: <https://medium.com/veri-madencili%C4%9Fi/veri-madencili%C4%9Finin-tarihi-ve-kullan%C4%B1ld%C4%B1%C4%9F%C4%B1-alanlar-19a1896c1505>. [Erişildi: 5 Aralık 2022].
- [37] H. Meng, J. Song, W. Hong and S. Li, "Knowledge Discovery from Categorical Data Based on Structured Partial Ordered Attribute Diagram," in *2018 Eighth International Conference on Instrumentation & Measurement, Computer, Communication and Control (IMCCC)*, China, 2018.
- [38] Oracle, «Data Mining Concepts,» 2008. [Çevrimiçi]. Available: https://docs.oracle.com/cd/B28359_01/datamine.111/b28129/process.htm#DMCON002. [Erişildi: 03 Aralık 2022].
- [39] U. M. Fayyad, "Data Mining and Knowledge Discovery in Databases: Implications for Scientific Databases," in *Ninth Internation Conferance on Scientific and Statistical Database Management, IEEE*, 1997.
- [40] G. Silahtaroglu, *Veri Madenciliği*, İstanbul: Papatya Yayınları, 2008.

- [41] A. Samuel, «Some Studies in Machine Learning Using the Game of Checkers,» *IBM Journal of Research and Development*, pp. 535-545, 1959.
- [42] S. Pattanayak, *Pro Deep Learning with TensorFlow: A Mathematical Approach to Advanced Artificial Intelligence in Python*, Bangalore, India: Apress, 2017, pp. 153-221.
- [43] Y. Wu and W. Qiu, "Facial Expression Recognition Based on Improved Deep Belief Networks," in *AIP Conference Proceedings*, 2017.
- [44] Anonim2, «What is Supervised Learning?,» IBM Cloud Education, [Çevrimiçi]. Available: <https://www.ibm.com/topics/supervised-learning>. [Erişildi: 19 Aralık 2022].
- [45] H. B. Demir, «MAKİNE ÖĞRENMESİ – MACHINE LEARNING,» 23 Şubat 2021. [Çevrimiçi]. Available: <https://mektebiendustri.com/index.php/2021/02/23/makine-ogrenmesi-machine-learning/>. [Erişildi: 30 Aralık 2022].
- [46] L. P. Kaelbling, M. L. Littman ve A. W. Moore, «Reinforcement Learning: A Survey,» *Journal of Artificial Intelligence Research*, no. 4, pp. 237-285, 1996.
- [47] V. Castelli, S. T. Hutchins, C.-S. Li ve J. J. E. Turek, *Modifying an Unreliable Training Set for Supervised Classification*, Armonk, NY (US) : International Business Machines Corporation, 2001.
- [48] Z. Liu, «Unsupervised Learning with Graph Theoretical Algorithms and Its Applications to Transcriptomic Data Analysis,» PhD. Dissertation, Anatomy and Genetics, Imperial College, London, England, 2018.
- [49] Y. M. Kızılkaya ve A. Oğuzlar, «Bazı Denetimli Öğrenme Algoritmalarının R Programlama Dili ile Kıyaslanması,» *Karadeniz*, cilt 37, pp. 90-98, 2018.
- [50] H. Akpınar, «Veri Tabanlarında Bilgi Keşfi ve Veri Madenciliği,» *İ. U. İşletme Fakültesi Dergisi*, cilt 29 (1), pp. 1-22, 2000.
- [51] F. Höppner, F. Klawonn, R. Kruse ve T. Runkler, «Fuzzy Cluster Analysis: Methods for Classification, Data Analysis and Image Recognition,» *Chichester: John Wiley & Sons*, p. 8, 1999.

- [52] J. Han, J. Pei ve M. Kamber, Data mining: concepts and techniques, USA: Elsevier, 2011, p. 20.
- [53] M. S. Aldenderfer ve R. K. Blashfield, «Cluster Analysis,» Sage Publications, pp.16, California, 1984.
- [54] S. Na, L. Xumin and G. Yong, "Research on K-Means Clustering Algorithm: An Improved K-Means Clustering Algorithm," in *International Symposium on Intelligent Information Technology and Security Informatics, IITSI*, 2010.
- [55] G. Shmueli, P. C. Bruce, M. L. Stephens ve N. R. Patel, Data Mining for Business Analytics: Concepts, Techniques and Applications with JMP Pro, Canada: John Wiley & Sons, 2017.
- [56] B. S. Everitt ve T. Hothorn, An Introduction to Applied Multivariate Analysis with R, New York: Springer, 2011.
- [57] R. Shaikh, «Feature Selection Techniques in Machine Learning with Python,» 28 Ekim 2018. [Çevrimiçi]. Available: <https://towardsdatascience.com/feature-selection-techniques-in-machine-learning-with-python-f24e7da3f36e>. [Erişildi: 02 Ocak 2023].
- [58] H. Budak, «Özellik Seçim Yöntemleri ve Yeni Bir Yaklaşım,» *Süleyman Demirel Üniversitesi Fen Bilimleri Enstitüsü Dergisi*, cilt 22, pp. 21-31, 2018.
- [59] T. Caliński ve J. Harabasz, «A Dendrite Method for Cluster Analysis,» *Communications in Statistics theory and Methods*, cilt 3, no. 1, pp. 1-27, 1974.
- [60] D. L. Davies ve D. W. Bouldin, «A Cluster Separation Measure,» *IEEE Transactions on Pattern Analysis and Machine Intelligence*, no. 2, pp. 224-227, 1979.
- [61] D. J. Ketchen ve C. L. Shook, «The Application of Cluster Analysis in Strategic Management Research: An Analysis and Critique,» *Strategic Management Journal*, cilt 17, no. 6, pp. 441-458, 1996.
- [62] J. MacQueen, "Some Methods for Classification and Analysis of Multivariate Observations," in *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, California, 1967, pp. 281-297.

- [63] L. Kaufman ve P. J. Rosseeauw, *Finding Groups in Data: An Introduction to Cluster Analysis*, New York: John Wiley and Sons Inc., 1990.
- [64] G. Foresti, C. Piciarelli ve C. Micheloni, «Kernel-based clustering,» *IET Digital Library*, cilt 19, no. 42, pp. 113-114, 2013.
- [65] Y. Thakare ve S. Bagal, «Performance Evaluation of K-means Clustering Algorithm with Various Distance Metrics,» *IEEE Trans*, cilt 110, no. 11, pp. 12-16, 2015.
- [66] T. Zhang, R. Ramakrishnan and M. Livny, "BIRCH: An Efficient Data Clustering Method for Very Large Databases," in *ACM SIGMOD International Conference on Management of Data*, Montreal, Quebec, Canada, 1996.
- [67] W. Day ve H. Edelsbrunner, «Efficient Algorithms for Agglomerative Gaussian Mixture Model Clustering Methods,» *Journal of Classification*, cilt 1, no. 1, p. 7–24, 1984.
- [68] R. Xu ve D. C. Wunsch, *Clustering*, New Jersey: Wiley-IEEE Press, 2008.
- [69] L. Bocci and I. Mingo, "Clustering Large Data Set: An Applied Comparative Study," in *Advanced Statistical Methods for the Analysis of Large Data-Sets*, Berlin, Springer, 2012, pp. 3-12.
- [70] W. A. Aldana, *Data Mining Industry: Emerging Trends and New Opportunities*, Massachusetts: Massachusetts Institute of Technology, 2000.