YILDIZ TEKNİK ÜNİVERSİTESİ
KİMYA-METALÜRJİ FAKÜLTESİ
MATEMATİK MÜHENDİSLİĞİ BÖLÜMÜ

BİTİRME TEZİ

LOJİSTİK SEKTÖRÜNDE MÜŞTERİ DAVRANIŞLARININ MAKİNE ÖĞRENMESİ İLE SEGMENTASYONU

Tez Danışmanı: Doç. Dr. Bayram Ali İBRAHİMOĞLU

21052092, Sedanur DURĞUN

İstanbul, 2026

Ó Bu tezin bütün hakları Yıldız Teknik Üniversitesi Matematik Mühendisliği Bölümü'ne aittir.

İÇİNDEKİLER

ÖNSÖZ vii
ÖZET viii
ABSTRACT ix
1. GİRİŞ 1
1.1 Tedarik Zinciri Analitiğinin Önemi 1
1.2 Çalışmanın Amacı ve Kapsamı 1
1.3 Veri Setinin Tanıtımı (DataCo Supply Chain Verisi) 1
2. LİTERATÜR VE TEORİK TEMELLER 2
2.1 Müşteri İlişkileri Yönetimi (CRM) ve Segmentasyon 2
2.2 RFM Analizi: Recency, Frequency, Monetary 2
2.3 Tedarik Zinciri ve Lojistik Performans Göstergeleri 3
2.4 RFM Modelinin Lojistik Metriklerle Genişletilmesi (RFM-L Modeli) 3
3. MAKİNE ÖĞRENMESİ VE KÜMELEME ALGORİTMALARI 5
3.1 Denetimsiz Öğrenme (Unsupervised Learning) Kavramı 5
3.2 K-Means Kümeleme Algoritması 6
3.2.1 Algoritmanın Çalışma Prensibi 6
3.2.2 Optimizasyon: Dirsek (Elbow) Yöntemi ve WCSS 7
3.3 Hiyerarşik Kümeleme (Agglomerative Clustering) 8
3.4 Boyut İndirgeme: Temel Bileşenler Analizi (PCA) 9
4. VERİ ÖN İŞLEME VE METODOLOJİ 10
4.1 Veri Temizleme ve Eksik Değer Analizi 10
4.2 Özellik Mühendisliği (Feature Engineering) 10
4.2.1 Lojistik Gecikme (Shipping Delay) Metriğinin Türetilmesi 11
4.3 Aykırı Değer Analizi (Interquartile Range- IQR Yöntemi) 11
4.4 Veri Normalizasyonu ve Ölçeklendirme (StandartScaler) 12
5. UYGULAMA VE ANALİZ BULGULARI 14
5.1 Optimum Küme Sayısının Belirlenmesi 14
5.2 K-Means Modeli Uygulanması ve Segmentlerin Belirlenmesi 15
5.3 Model Performans Değerlendirmesi (Validasyon) 17
5.3.1 Silhoutte Skoru 17
5.3.2 Davies- Bouldin ve Calinski- Harabsz İndeksleri 18
5.4 PCA ile Segmentlerin Görselleştirilmesi 18
5.5 Segment Bazlı Stratejik Analiz ve Yorumlar 20

5.5.1 Sadık Müşteriler ve Lojistik Beklentileri 20
5.5.2 Kaybedilme Riski Olan Değerli Müşteriler (Lojistik Mağdurları) 20
5.5.3 Düşük Değerli ve Verimsiz Gruplar 21
6. SONUÇLAR VE ÖNERİLER 22
6.1 Elde Edilen Bulguların Değerlendirilmesi 22
6.2 Tedarik Zinciri Yöneticileri İçin Karar Destek Önerileri 22
6.3 Gelecek Çalışmalar İçin Yönlendirmeler 23
KAYNAKÇA 24
EKLER 30
..... 30
..... 32
ÖZGEÇMİŞ 35

ŞEKİLLER LİSTESİ

Şekil 5.1. Dirsek Metodu Grafiği 15
Şekil 5.2. Segment Bazlı İstatistikler 16
Şekil 5.3. Model Güven Değerleri 17
Şekil 5.4. PCA ile Müşteri Segmentasyonu Gösterimi 19

ÖNSÖZ

Bu çalışma, **Yıldız Teknik Üniversitesi Matematik Mühendisliği Bölümü Bitirme Tezi** projesi kapsamında hazırlanmıştır.

Veri bilimi ve makine öğrenmesi teknikleri, günümüzde yalnızca mühendislik alanlarında değil, küresel ticaretin ve ekonominin önemli noktalarından olan tedarik zinciri yönetimi ve lojistik sektöründe de devrim niteliğinde çözümler sunmaktadır. Bu çalışma ile, denetimsiz öğrenme yöntemlerinin en temel ve etkili algoritmalarından biri olan K-Means kümeleme modelinin sadece teorik temellerini incelemekle kalmayıp, DataCo tedarik zinciri veri seti üzerinden müşteri davranışları ve lojistik performansın entegrasyonu ile bu teknolojinin gerçek dünya problemlerine nasıl çözüm ürettiği gösterilmiştir. Projenin, hem müşteri segmentasyonu analizlerini lojistik parametrelerle güçlendirmesi hem de veriye dayalı stratejik karar destek sistemlerine yönelik ileri düzey çalışmalara bir temel oluşturması hedeflenmiştir.

Bu proje konusunun seçilme nedeni; modern ticaret dünyasında büyük veri kullanımının yaygınlığına rağmen, müşteri sadakati ile operasyonel lojistik süreçlerin bir etkileşim içinde, sistematik bir matematiksel ve istatistiksel yaklaşımla ele alınması gerekliliğidir. Özellikle geleneksel RFM (Recency, Frequency, Monetary) modeline lojistik performans metriklerinin eklenmesiyle ortaya çıkan yeni perspektife duyulan merak, konunun teorik yönüyle pratik uygulamasının bu proje kapsamında birleştirilmesini sağlamıştır.

Projenin hazırlanma sürecinde elde edilen deneyimler, sadece kümeleme algoritmalarının ve boyut indirgeme (PCA) tekniklerinin daha iyi anlaşılmasını sağlamakla kalmamış; aynı zamanda büyük veri setlerinin temizlenmesi, özellik mühendisliği, model optimizasyonu ve akademik bir metnin etik çerçevede yazım süreci hakkında şahsıma önemli kazanımlar sunmuştur.

Bu projenin hazırlanma sürecinde bilgilerini benden esirgemeyen, yönlendirmeleriyle çalışmama vizyon katan danışman hocam **Doç. Dr. Bayram Ali İBRAHİMOĞLU**'na, ayrıca lisans eğitimim boyunca akademik altyapımın oluşmasında emeği geçen tüm öğretim üyelerine teşekkürlerimi sunarım.

Bunun yanında, eğitim hayatım boyunca maddi ve manevi desteklerini esirgemeyen, her daim yanımda olan aileme ve sevgili arkadaşlarıma bana gösterdikleri sevgi, sabır ve güvenlerinden dolayı en içten teşekkürlerimi borç bilirim.

ÖZET

Bu çalışmanın amacı, tedarik zinciri yönetiminde denetimsiz makine öğrenmesi tekniklerini kullanarak lojistik performans odaklı, dinamik bir müşteri segmentasyonu modeli geliştirmektir.

Projede öncelikle, pazarlama analitiğinin temelini oluşturan geleneksel RFM (Recency, Frequency, Monetary) modelinin teorik altyapısı incelenmiş; bu modele lojistik operasyon süreçlerinin bir göstergesi olan "Shipping Delay" (Sevkiyat Gecikmesi) değişkeni eklenerek model genişletilmiştir. Veri madenciliği süreçleri kapsamında, büyük veri setleri üzerinde aykırı değer analizi (IQR), veri normalizasyonu ve boyut indirgeme (PCA) teknikleri sistematik bir matematiksel yaklaşımla uygulanmıştır.

Kümeleme aşamasında K-Means algoritması kullanılarak müşteriler karakteristik özelliklerine göre ayrıştırılmış; optimum küme sayısı "Dirsek Yöntemi" ile belirlenmiştir. Geliştirilen modelin başarısı; Silhouette Skoru, Davies-Bouldin Endeksi ve Calinski-Harabasz Endeksi gibi validasyon metrikleriyle matematiksel olarak doğrulanmıştır.

Teorik çerçeve, DataCo Supply Chain veri seti üzerinden yapılan uygulama ile desteklenmiş; algoritmanın çok boyutlu verileri anlamlandırmadaki etkinliği görselleştirme teknikleriyle sunulmuştur. Elde edilen sonuçlar, kümeleme analizinin ve istatistiksel modellemenin, karmaşık tedarik zinciri problemlerini çözmede, müşteri sadakatini ölçmede ve operasyonel verimliliği artırmada ne kadar güçlü bir araç olduğunu ortaya koymaktadır.

Anahtar Kelimeler: Matematiksel Modelleme, Makine Öğrenmesi, K-Ortalamalar Kümeleme, RFM Analizi, Tedarik Zinciri Yönetimi, Lojistik Optimizasyonu, Veri Madenciliği.

ABSTRACT

The objective of the study is to develop a dynamic customer segmentation model focused on logistics performance by utilizing unsupervised machine learning techniques in supply chain management.

In the initial phase of the study, the theoretical framework of the traditional RFM (Recency, Frequency, Monetary) model, which forms the basis of marketing analytics, was examined. This model was subsequently extended by integrating the "Shipping Delay" variable as a key indicator of operational logistics processes. Within the scope of data mining procedures, outlier analysis (IQR), data normalization, and dimensionality reduction (PCA) techniques were applied through a systematic mathematical approach.

During the clustering stage, customers were segmented according to their characteristic behaviors using the K-Means algorithm, where the optimal number of clusters was determined via the "Elbow Method." The performance and reliability of the developed model were mathematically validated using metrics such as the Silhouette Score, Davies-Bouldin Index, and Calinski-Harabasz Index.

The theoretical framework was supported by an empirical application using the DataCo Supply Chain dataset, and the effectiveness of the algorithm in interpreting high-dimensional data was demonstrated through advanced visualization techniques. The results reveal that clustering analysis and statistical modeling are powerful instruments for addressing complex supply chain challenges, measuring customer loyalty, and enhancing operational efficiency.

Keywords: Mathematical Modeling, Machine Learning, K-Means Clustering, RFM Analysis, Supply Chain Management, Logistics Optimization, Data Mining.

1. GİRİŞ

Günümüzün küresel ekonomi düzeninde, dijitalleşme ve veri üretim hızındaki devasa artış, işletmeleri geleneksel karar alma mekanizmalarından uzaklaştırarak veri odaklı stratejiler geliştirmeye zorlamaktadır. Özellikle tedarik zinciri yönetimi ve lojistik operasyonlar, bir işletmenin maliyet yapısını ve müşteri memnuniyetini doğrudan etkileyen en kritik süreçler haline gelmiştir [1]. Bu bağlamda, "büyük veri analitiği" ve "makine öğrenmesi" teknikleri, karmaşık veri yığınlarından anlamlı paternler çıkarılmasına ve operasyonel verimliliğin artırılmasına imkan tanımaktadır [2].

1.1 Tedarik Zinciri Analitiğinin Önemi

Tedarik zinciri süreçleri; ham madde tedarikinden üretim aşamasına, depolamadan son kullanıcıya ulaşıma kadar uzanan heterojen bir yapıya sahiptir. Bu süreçte oluşan verilerin analizi, sadece geçmiş performansı ölçmekle kalmaz, aynı zamanda gelecekteki talepleri ve olası darboğazları öngörmeyi sağlar. Matematiksel modelleme teknikleri, tedarik zincirindeki bu belirsizlikleri yönetilebilir kılan en güçlü araçtır. Modern lojistik yönetiminde, müşteriyi sadece satın alma gücüyle değil, aynı zamanda lojistik süreçlerden aldığı hizmet kalitesiyle değerlendirmek stratejik bir zorunluluk haline gelmiştir [3].

1.2 Çalışmanın Amacı ve Kapsamı

Bu çalışmanın temel amacı, pazarlama analitiği literatüründe yer alan geleneksel RFM (Recency, Frequency, Monetary) modelini, lojistik performans metrikleri ile zenginleştirerek daha kapsamlı bir müşteri segmentasyonu modeli oluşturmaktır [4]. Klasik modeller müşterinin sadece harcama tutarına ve sıklığına odaklanırken; bu çalışmada modele "Lojistik Gecikme" (Shipping Delay) faktörü entegre edilerek literatürdeki hibrit model arayışlarına katkı sunulması hedeflenmiştir.

Çalışma kapsamında, denetimsiz öğrenme yöntemlerinden **K-Means (K-Ortalamalar)** kümeleme algoritması kullanılarak müşteriler karakteristik benzerliklerine göre gruplandırılmıştır. K-Means, büyük veri setlerinin kümelenmesinde hızı ve etkinliği nedeniyle sıkça tercih edilen bir matematiksel yaklaşımdır [5]. Modelin doğrulanması sürecinde ise Silhouette Skoru ve Dirsek Yöntemi gibi metrikler kullanılarak optimizasyon sağlanmıştır. Ayrıca, çok boyutlu verilerin görselleştirilmesi ve varyans analizi için Temel Bileşenler Analizi (PCA) tekniğine başvurulmuştur.

1.3 Veri Setinin Tanıtımı (DataCo Supply Chain Verisi)

Uygulama aşamasında, Constante ve arkadaşları tarafından literatüre sunulan ve gerçek dünya operasyonel verilerini içeren "DataCo Supply Chain Dataset" kullanılmıştır [6]. Bu veri seti; finansal veriler, sipariş detayları ve sevkiyat bilgileri gibi 50'den fazla değişkeni içermektedir. Analiz sürecinde, bu geniş veri havuzundan müşteri davranışlarını ve lojistik performansı en iyi temsil eden değişkenler seçilerek temizlenmiş ve özellik mühendisliği süreçlerinden geçirilmiştir.

2. LİTERATÜR VE TEORİK TEMELLER

Bu bölümde, çalışmanın temelini oluşturan Müşteri İlişkileri Yönetimi (CRM) ve geleneksel RFM (Recency, Frequency, Monetary) analizinin literatürdeki tarihsel gelişimi ele alınmıştır. Projenin özgün değerini oluşturan lojistik performans göstergelerinin bu modele entegrasyonu kuramsal çerçevede incelendikten sonra, veri madenciliği süreçlerinde gerek duyulacak temel istatistiksel kavramlar sunulmuştur. Bölümün devamında, müşteri davranışlarını gruplandırmak için kullanılan denetimsiz öğrenme yöntemleri ile kümeleme analizinin matematiksel dayanağını oluşturan K-Means algoritmasının ve boyut indirgeme tekniklerinin teorik altyapısı detaylıca incelenmiştir..

2.1 Müşteri İlişkileri Yönetimi (CRM) ve Segmentasyon

Modern işletme yönetiminde Müşteri İlişkileri Yönetimi (CRM), sadece teknolojik bir altyapı veya veritabanı çözümü değil, stratejik bir iş felsefesi olarak konumlandırılmaktadır [7]. Geleneksel pazarlamanın temelini oluşturan ve geniş kitlelere tek tip mesaj iletmeyi hedefleyen "kitle pazarlaması" paradigması, yerini her bir müşterinin bireysel ihtiyaçlarını merkeze alan "ilişkisel pazarlama" anlayışına bırakmıştır [8]. CRM'in nihai amacı, müşteri etkileşimlerini analiz ederek davranışsal kalıpları öngörmek ve "Müşteri Yaşam Boyu Değeri"ni (Customer Lifetime Value - CLV) maksimize ederek işletme karlılığını sürdürülebilir kılmaktır [9].

Bu stratejik yaklaşımın merkezinde ise segmentasyon süreci yer almaktadır. Literatürde müşteri segmentasyonu, heterojen yapıdaki bir pazarın; benzer satın alma alışkanlıklarına, ihtiyaçlara veya karakteristik özelliklere sahip homojen alt gruplara ayrıştırılması işlemi olarak tanımlanır [10]. Etkin bir segmentasyon, işletmelerin kısıtlı kaynaklarını verimsiz alanlar yerine, yüksek sadakat veya karlılık potansiyeli sunan segmentlere yönlendirmesine olanak tanır. Günümüzde veri bilimi tekniklerinin gelişimiyle birlikte; yaş ve cinsiyet gibi durağan demografik verilerin yerini, işlem geçmişi ve servis kullanım sıklığı gibi dinamik "davranışsal veriler" almıştır [11]. Bu çalışmada uygulanan kümeleme analizi, verideki doğrusal olmayan karmaşık ilişkileri keşfederek, geleneksel yöntemlerle tespit edilemeyecek stratejik müşteri gruplarının tanımlanmasını sağlamaktadır.

2.2 RFM Analizi: Recency, Frequency, Monetary

Müşteri sadakatini ve karlılığını ölçümlemede literatürde en çok atıf alan modellerden biri RFM (Recency, Frequency, Monetary) analizidir. İlk olarak Hughes (1994) tarafından popülerleştirilen bu model, müşterilerin geçmiş işlem verilerini kullanarak gelecekteki satın alma olasılıklarını tahmin eden sayısal bir skorlama sistemidir [4]. Modelin üç temel bileşeni şu şekilde kuramsallaştırılmıştır:

- **Recency (Yenilik):** Müşterinin son etkileşiminden bu yana geçen süreyi ifade eder. Pazarlama literatürü, yakın zamanda işlem yapmış bir müşterinin, uzun süredir aktif olmayan bir müşteriye kıyasla yeni bir pazarlama teklifine yanıt verme olasılığının daha yüksek olduğunu kanıtlamaktadır [12].

- **Frequency (Sıklık):** Belirli bir zaman dilimi içerisinde gerçekleştirilen toplam işlem sayısıdır. Bu metrik, müşteri sadakatinin ve markaya olan bağlılığın doğrudan bir göstergesi olarak kabul edilir [13].

- **Monetary (Parasal Değer):** Müşterinin toplamda yarattığı ekonomik hacmi temsil eder. İşletmelerin finansal sürdürülebilirliği açısından en kritik gösterge olup, yüksek değerli müşterilerin (VIP) tespitinde birincil rol oynar [14].

RFM modelinin en güçlü yönü, veri analitiği sonuçlarının iş birimleri tarafından kolayca yorumlanabilmesidir. Ancak, modelin işlem odaklı yapısı, özellikle lojistik ve operasyonel süreçlerin yoğun olduğu sektörlerde müşteri memnuniyetini etkileyen dışsal faktörleri analiz dışı bırakabilmektedir [15].

2.3 Tedarik Zinciri ve Lojistik Performans Göstergeleri

Tedarik zinciri yönetimi, ürünün başlangıç noktasından nihai tüketiciye kadar olan fiziksel ve bilgi akışının bütünsel optimizasyonunu hedefler [16]. E-ticaretin küresel ölçekte yaygınlaşmasıyla birlikte, fiziksel ürün kalitesi kadar, teslimat hızı ve güvenilirliği de birincil rekabet avantajı haline gelmiştir. Lojistik performans, müşteri tatminini (customer satisfaction) ve marka imajını doğrudan etkileyen operasyonel bir dışsallıktır [1].

Akademik çalışmalarda lojistik başarı genellikle "Hizmet Seviyesi" (Service Level) parametreleri üzerinden değerlendirilir. Bu tez çalışmasında lojistik performansı ölçümlemek adına "Gerçekleşen Gönderim Süresi" ve "Planlanan Gönderim Süresi" arasındaki korelasyon temel alınmıştır. Bir lojistik sistemin etkinliği, sadece "hız" ile değil, aynı zamanda vaat edilen tarihe uyum (teslimat güvenilirliği) ile ölçülür [17]. Planlanan süreden sapmalar, tedarik zinciri ağındaki verimsizliklerin, stok yönetimi hatalarının veya taşımacı performansındaki aksaklıkların bir göstergesidir. Özellikle sınır ötesi operasyonlarda gümrükleme süreçleri ve taşıma modları (intermodal, hava yolu vb.) bu performans çıktıları üzerinde belirleyici unsurlardır [18].

2.4 RFM Modelinin Lojistik Metriklerle Genişletilmesi (RFM-L Modeli)

Geleneksel RFM analizi, pazarlama perspektifinden güçlü bir segmentasyon sunsa da, operasyonel aksaklıkların müşteri kaybı (churn) üzerindeki etkisini açıklamakta yetersiz kalmaktadır. Bu eksikliği gidermek adına literatürde, standart RFM boyutlarına sektörel değişkenlerin eklendiği "Genişletilmiş RFM Modelleri" geliştirilmektedir [19]. Bu çalışmada, lojistik performansın müşteri davranışına etkisini ölçmek amacıyla literatüre önerilen "RFM-L" (Logistics) yaklaşımı benimsenmiştir.

Modelin "L" bileşeni, "Sevkiyat Gecikmesi" (Shipping Delay) değişkeni ile temsil edilmektedir. Bu değişken, gerçekleşen teslimat süresinden planlanan sürenin çıkarılmasıyla elde edilen "sapma değerini" ifade eder. Bu genişletmenin teorik dayanağı, lojistik hizmet kalitesinin müşteri sadakati üzerindeki dolaylı etkisidir: Yüksek harcama kapasitesine (Monetary) ve alışveriş sıklığına (Frequency) sahip bir

müşterinin, sürekli yaşanan lojistik gecikmeler nedeniyle markadan uzaklaşma riski (churn risk) literatürde vurgulanmaktadır [20]. RFM-L modeli sayesinde, sadece finansal getirisi yüksek olanlar değil, aynı zamanda operasyonel süreçlerde mağduriyet yaşayan "hassas" müşteri grupları da tespit edilebilmektedir. Bu yaklaşım, CRM stratejilerini lojistik mükemmellik ile birleştirerek daha kapsayıcı bir müşteri yönetim modeli sunmaktadır [21].

3. MAKİNE ÖĞRENMESİ VE KÜMELEME ALGORİTMALARI

Veri bilimi ve yapay zeka teknolojilerindeki paradigma değişimi, işletmelerin sahip olduğu büyük ölçekli ve karmaşık veri yığınlarından stratejik değer taşıyan bilgilerin çıkarılmasını mümkün kılmıştır [22]. Geleneksel istatistiksel yöntemlerin, değişken sayısının arttığı ve veriler arasındaki doğrusal olmayan ilişkilerin karmaşıklaştığı durumlarda yetersiz kalması, makine öğrenmesi algoritmalarına olan ihtiyacı doğurmuştur [23]. Makine öğrenmesi; açıkça programlanmaya gerek duyulmaksızın, veriden otomatik olarak öğrenen, deneyim yoluyla performansını geliştiren ve belirli görevleri yerine getirmek için matematiksel modeller inşa eden sistemlerin genel adıdır [24]. Bu çalışma kapsamında tercih edilen algoritmalar, müşteri davranış örüntülerini ve lojistik performans metriklerini çok boyutlu bir düzlemde analiz ederek, verinin kendi içindeki doğal hiyerarşisini ve gruplandırmalarını insan müdahalesi olmaksızın ortaya çıkarmayı hedeflemektedir.

Bu bölümde, çalışmanın metodolojik omurgasını oluşturan denetimsiz öğrenme disiplini üzerinde durulacaktır. İlk olarak, veri setindeki örtük yapıların keşfedilmesinde kullanılan kümeleme (clustering) teorisi ele alınacaktır. Ardından, çalışmanın ana analiz motoru olan K-Means algoritması, bu algoritmanın optimizasyon süreçleri ve doğrulama amacıyla başvurulan hiyerarşik kümeleme yöntemleri detaylandırılacaktır. Ayrıca, çok boyutlu veri setlerinin (RFM-L gibi) anlamlandırılması ve görselleştirilmesinde kritik bir öneme sahip olan boyut indirgeme tekniklerinden Temel Bileşenler Analizi'nin (Principal Component Analysis - PCA) kuramsal çerçevesi sunulacaktır. Bu teorik zemin, uygulama safhasında elde edilen sonuçların bilimsel geçerliliğini ve yorumlanabilirliğini destekleyen bir dayanak oluşturacaktır [25].

3.1 Denetimsiz Öğrenme (Unsupervised Learning) Kavramı

Makine öğrenmesi literatürü, öğrenme biçimlerine göre genellikle üç ana disipline ayrılmaktadır: denetimli (supervised), denetimsiz (unsupervised) ve pekiştirmeli (reinforcement) öğrenme [26]. Denetimsiz öğrenme; veri setinde önceden tanımlanmış sınıf etiketlerinin (labels) veya hedef çıktı değerlerinin bulunmadığı senaryolarda, verinin karakteristik özelliklerini, benzerliklerini ve gizli kalıplarını keşfetmek amacıyla kullanılan bir yöntemdir [27]. Denetimli öğrenmenin aksine, burada bir "öğretici" veya "doğru cevap anahtarı" bulunmaz; algoritma, girdi verilerinin öznitelik uzayındaki (feature space) dağılımlarını inceleyerek kendi kendine yapısal bir düzen oluşturur [28].

Müşteri segmentasyonu ve pazar analizi çalışmaları, denetimsiz öğrenmenin akademik ve endüstriyel alandaki en köklü uygulama alanlarından biridir [29]. İşletmelerin sahip olduğu ham işlem kayıtlarında, hangi müşterinin hangi stratejik kategoriye (örneğin; "sadık", "riskli" veya "potansiyeli yüksek") ait olduğu bilgisi önsel olarak mevcut değildir. Denetimsiz öğrenme algoritmaları, bu çalışmada temel alınan RFM (Recency, Frequency, Monetary) ve lojistik (L) metrikleri arasındaki karmaşık matematiksel ilişkileri ve mesafeleri hesaplayarak, benzer davranışsal profil sergileyen gözlemleri aynı kümeye dahil eder [30]. Bu süreç, sadece verinin özetlenmesini sağlamakla kalmaz, aynı zamanda yüksek boyutlu ve gürültülü verilerin stratejik karar destek mekanizmalarına dönüştürülmesini sağlayan bir keşifsel veri analizi süreci işlevi görür [31].

3.2 K-Means Kümeleme Algoritması

Gözetimsiz öğrenme (unsupervised learning) teknikleri arasında en yaygın kullanılan ve literatürde en köklü geçmişe sahip yöntemlerden biri K-Means kümeleme algoritmasıdır. İlk olarak 1967 yılında James MacQueen tarafından önerilen bu algoritma, veri setindeki benzer özelliklere sahip gözlemleri birbirinden ayırt ederek belirli sayıda gruba (kümeye) ayırmayı hedefler [32]. Temel felsefesi, veri noktaları ile bağlı bulundukları küme merkezleri (centroid) arasındaki varyansı minimize etmektir. K-Means, büyük veri setlerinde bile hesaplama maliyetinin düşük olması ve uygulama kolaylığı nedeniyle biyoinformatikten pazarlamaya, görüntü işlemeden anomali tespitine kadar geniş bir yelpazede tercih edilmektedir [5]. Algoritmanın temel amacı, veri seti içerisindeki her bir x_i gözlemini, en yakın merkez noktasına sahip olan $S = \{S_1, S_2, \ldots, S_k\}$ kümelerinden birine dahil ederek toplam küme içi kareler toplamını en aza indirmektir.

3.2.1 Algoritmanın Çalışma Prensibi

K-Means algoritması, iteratif (tekrarlayan) bir mantıkla çalışan, "bölümleyici" (partitioning) tabanlı bir algoritmadır. Süreç, kullanıcı tarafından önceden belirlenen bir k parametresi (küme sayısı) ile başlar. Algoritmanın işleyişi genel olarak şu dört ana aşamadan oluşmaktadır [29] :

1. **Başlatma (Initialization):** Veri setinden rastgele k adet nokta seçilir ve bu noktalar başlangıç küme merkezleri (centroidler) olarak tanımlanır.
2. **Atama (Assignment):** Veri setindeki her bir gözlem, en yakın küme merkezine atanır. Yakınlık ölçütü olarak genellikle Öklid mesafesi (Euclidean Distance) kullanılır. İki nokta (x ve y) arasındaki Öklid mesafesi şu formülle hesaplanır [33]:

$$d(x,y) = \sqrt{\sum_{i=1}^{n} (x_i - y_i)^2}$$

3. **Güncelleme (Updating):** Tüm noktalar bir kümeye atandıktan sonra, her kümenin merkezi, o kümeye ait olan tüm noktaların aritmetik ortalaması alınarak yeniden hesaplanır. Yeni merkez noktası (μ_j), S_j kümesindeki n_j adet noktanın ortalamasıdır:

$$\mu_j = \frac{1}{|S_j|} \sum_{x_i \in S_j} x_i$$

4. **Yakınsama (Convergence):** Atama ve güncelleme adımları, küme merkezleri artık değişmeyene kadar veya önceden belirlenen maksimum iterasyon sayısına ulaşılana kadar devam eder.

Algoritma, yerel minimuma takılma riskine sahip olsa da, farklı başlangıç noktalarıyla (K-Means++ gibi yöntemlerle) bu sorunun önüne geçilebilmektedir [34]. Algoritmanın

nihai başarısı, kümelerin kendi içindeki homojenliğinin yüksek, kümeler arası heterojenliğin ise maksimum olmasına bağlıdır.

3.2.2 Optimizasyon: Dirsek (Elbow) Yöntemi ve WCSS

K-Means algoritmasının en kritik dezavantajlarından biri, küme sayısı olan k değerinin kullanıcı tarafından dışsal olarak belirlenmesinin gerekliliğidir. Yanlış bir k seçimi, verinin doğal yapısının yanlış yorumlanmasına yol açabilir. Bu karmaşıklığı çözmek ve optimal küme sayısını matematiksel bir temele oturtmak için literatürde en sık başvurulan yöntem "Dirsek Yöntemi" (Elbow Method) ve bu yöntemin temelini oluşturan WCSS (Within-Cluster Sum of Squares) değeridir [35].

WCSS (Küme İçi Kareler Toplamı): WCSS, her bir veri noktasının kendi küme merkezine olan uzaklıklarının karelerinin toplamıdır. Matematiksel olarak şu şekilde ifade edilir:

$$WCSS = \sum_{j=1}^{k} \sum_{x_i \in S_j} \left\| x_i - \mu_j \right\|^2$$

Burada k küme sayısını, x_i veri noktasını ve μ_j ise j. kümenin merkezini temsil eder. WCSS değeri, kümelemenin ne kadar "sıkı" veya "yoğun" olduğunu gösteren bir performans metriğidir. Küme sayısı arttıkça, noktalar merkezlere daha yakın olacağı için WCSS değeri doğal olarak azalır. k sayısı veri sayısına eşit olduğunda WCSS sıfıra iner; ancak bu durum modelin genelleme yeteneğini yok eder (overfitting) [36].

Dirsek (Elbow) Yöntemi: Optimal k değerini bulmak için, farklı k değerlerine (örneğin 1'den 10'a kadar) karşılık gelen WCSS değerleri hesaplanır ve bir grafik üzerinde görselleştirilir. Grafik incelendiğinde, WCSS değerindeki düşüşün en keskin olduğu noktadan sonra eğilimin yavaşladığı ve grafiğin bir "dirsek" şekli oluşturduğu görülür [8]. Bu kırılma noktası (inflection point), verideki varyansın açıklanması ile küme sayısı arasındaki en ideal dengeyi temsil eder. Bu noktadan sonra küme sayısını artırmak, WCSS değerinde anlamlı bir iyileşme sağlamaz; bu nedenle dirsek noktasındaki k değeri, veri seti için en uygun küme sayısı olarak kabul edilir [37].

3.3 Hiyerarşik Kümeleme (Agglomerative Clustering)

Hiyerarşik kümeleme, veri noktaları arasında bir hiyerarşi oluşturarak gruplandırma yapan ve sonuçları genellikle bir ağaç yapısı (dendrogram) ile sunan bir denetimsiz öğrenme yöntemidir. K-Means algoritmasından farklı olarak, bu yöntemde başlangıçta küme sayısının (k) belirlenmesine gerek duyulmaz; bunun yerine veri noktalarının birbirine olan yakınlıkları üzerinden kademeli bir birleştirme veya bölme işlemi gerçekleştirilir [25]. Literatürde hiyerarşik kümeleme iki temel yaklaşıma ayrılır: "Tümevarımcı" (Agglomerative) ve "Tümdengelimci" (Divisive). Bu çalışmada en sık kullanılan yöntem olan ve aşağıdan yukarıya doğru çalışan Birleştirici (Agglomerative) yaklaşım temel alınmıştır.

Çalışma Mekanizması: Birleştirici hiyerarşik kümeleme süreci şu adımları izler [38]:

Başlangıçta her bir veri noktası tek başına bağımsız bir küme olarak kabul edilir.

Birbirine en yakın olan (en yüksek benzerliğe sahip) iki küme belirlenir ve birleştirilir.

Bu işlem, tüm veri noktaları tek bir büyük kümede toplanana kadar iteratif olarak devam eder.

Bağlantı Yöntemleri (Linkage Methods): Kümeler arasındaki mesafenin nasıl hesaplanacağı, algoritmanın sonuçlarını doğrudan etkiler. Literatürde kabul görmüş temel bağlantı yöntemleri şunlardır [39]:

Single Linkage: İki küme arasındaki mesafeyi, kümelerdeki birbirine en yakın noktalar arasındaki mesafe olarak tanımlar.

Complete Linkage: İki küme arasındaki mesafeyi, kümelerdeki birbirine en uzak noktalar arasındaki mesafe olarak alır.

Average Linkage: Tüm noktalar arasındaki mesafelerin ortalamasını hesaplar.

Ward’s Method: Birleştirme işlemi sonucunda oluşacak küme içi varyans artışını minimize etmeyi hedefler. Bu yöntem, genellikle daha homojen ve dengeli kümeler oluşturduğu için tercih edilmektedir [40].

Dendrogram ve Karar Süreci: Hiyerarşik kümelemenin en büyük avantajı, "Dendrogram" adı verilen görselleştirme aracıdır. Dendrogram, verilerin hangi aşamada ve hangi mesafe eşiğinde birleştiğini gösteren bir diyagramdır. Araştırmacı, dendrogram üzerindeki dikey çizgilerin uzunluklarını inceleyerek verinin doğal yapısına en uygun küme sayısına karar verebilir [41]. Ancak, bu yöntemin hesaplama karmaşıklığı $O(n^2)$ veya $O(n^3)$ olduğu için çok büyük veri setlerinde K-Means'e göre daha yavaş çalışmaktadır [42].

3.4 Boyut İndirgeme: Temel Bileşenler Analizi (PCA)

Modern veri setleri, özellikle RFM ve lojistik metrikler gibi çok sayıda değişken içerdiğinde "Boyutun Laneti" (Curse of Dimensionality) olarak bilinen sorunla karşılaşılmaktadır. Değişken sayısı arttıkça verinin görselleştirilmesi zorlaşmakta ve modellerin hesaplama maliyeti artmaktadır. Temel Bileşenler Analizi (Principal Component Analysis - PCA), verideki bilgi kaybını minimumda tutarak değişken sayısını azaltan ve literatürde en yaygın kullanılan boyut indirgeme tekniğidir [43]. İlk olarak Karl Pearson (1901) tarafından önerilen bu yöntem, verideki varyansı (bilgiyi) en iyi temsil eden yeni ve birbirine dik (orthogonal) değişkenler türetir.

Matematiksel Süreç: PCA'nın işleyişi şu temel aşamalardan oluşur [44]:

1. **Standartlaştırma:** Farklı ölçeklerdeki değişkenlerin (örneğin; para birimi ve gün sayısı) analize eşit etkide bulunması için veriler standart normal dağılıma ($z - score$) dönüştürülür.

2. **Kovaryans Matrisi Hesaplama:** Değişkenlerin birbirleriyle olan ilişkilerini anlamak için kovaryans matrisi oluşturulur.

3. **Özdeğer ve Özvektör Hesabı:** Matris üzerinden özdeğerler (eigenvalues) ve özvektörler (eigenvectors) hesaplanarak verideki en yüksek varyansın hangi yönlerde olduğu belirlenir.

4. **Temel Bileşenlerin Seçimi:** En yüksek özdeğere sahip olan ilk n bileşen seçilir. "PC1" (Birinci Temel Bileşen) verideki en büyük varyansı açıklar, "PC2" ise PC1 ile ilişkisiz (dik) olup kalan varyansın en büyük kısmını açıklar [45].

Analizdeki Rolü: PCA, sadece boyut indirgemekle kalmaz, aynı zamanda veri setindeki çoklu doğrusallık (multicollinearity) problemini ortadan kaldırarak kümeleme algoritmalarının performansını artırır [46]. Özellikle RFM-L modelinde olduğu gibi birbiriyle ilişkili olabilecek (örneğin sıklık ve parasal değer) değişkenlerin etkisini dengelemek ve kümeleri 2 boyutlu veya 3 boyutlu düzlemde görselleştirebilmek için kritik bir öneme sahiptir [47]. "Scree Plot" adı verilen grafikler aracılığıyla, toplam varyansın ne kadarının korunacağı ve kaç bileşenle temsil edileceği bilimsel bir temel üzerinde belirlenir [48].

4. VERİ ÖN İŞLEME VE METODOLOJİ

Veri madenciliği ve makine öğrenmesi projelerinde elde edilecek sonuçların doğruluğu, büyük ölçüde kullanılan verinin kalitesine ve hazırlık süreçlerinin titizliğine bağlıdır. Literatürde "Çöp girerse, çöp çıkar" (Garbage In, Garbage Out) olarak bilinen ilke gereği, ham verilerin temizlenmeden modele dahil edilmesi, algoritmanın gürültülü verileri gerçek kalıplar olarak algılamasına ve hatalı stratejik kararlar alınmasına yol açmaktadır [49]. Bu çalışmada, "DataCo Supply Chain" veri seti üzerinde uygulanan metodolojik akış; veri temizleme, öznitelik mühendisliği, aykırı değer analizi ve standardizasyon adımlarını kapsamaktadır. Bu süreçte endüstri standardı olan CRISP-DM (Cross-Industry Standard Process for Data Mining) metodolojisi takip edilmiştir [50].

4.1 Veri Temizleme ve Eksik Değer Analizi

Analiz sürecinin ilk aşaması, verinin teknik olarak işlenebilir hale getirilmesidir. Veri seti yüklenirken kullanılan 'ISO-8859-1' kodlaması, uluslararası tedarik zinciri verilerindeki özel karakterlerin korunmasını sağlamıştır. Programatik süreçte hata payını azaltmak adına, sütun isimlerinde yer alan boşluklar ve özel karakterler alt tire (_) sembolü ile değiştirilerek standartlaştırılmıştır.

Veri temizleme aşamasının en kritik bileşeni eksik değer (missing value) yönetimidir. Veri setinde yapılan incelemeler sonucunda; 'Customer_Id', 'Order_Item_Total' ve lojistik performans hesaplamalarının temel girdisi olan 'Days_for_shipping_(real)' sütunlarında bazı boş kayıtlar olduğu gözlemlenmiştir. İstatistikte eksik veriler için ortalama atama (mean imputation) gibi yöntemler mevcut olsa da, bu çalışmada veri setinin hacminin oldukça geniş olması ve lojistik gibi hassas bir alanda yapay veri üretiminin modelin gerçekçiliğini bozabileceği düşüncesiyle, eksik veri içeren satırların veri setinden tamamen çıkarılması (listwise deletion) yoluna gidilmiştir [51]. Bu yaklaşım, analizin sadece gerçek ve doğrulanmış gözlemler üzerinden yürümesini garanti altına almıştır [52].

4.2 Özellik Mühendisliği (Feature Engineering)

Özellik mühendisliği, ham veri setinde bulunan değişkenlerin, makine öğrenmesi modellerinin öğrenme kapasitesini artırmak amacıyla yeni ve anlamlı özniteliklere dönüştürülmesi sürecidir. Ham veriler genellikle doğrudan bir stratejik çıktı üretmez; bu nedenle verideki gizli örüntülerin (patterns) algoritma tarafından yakalanabilmesi için verinin işletme mantığına göre yeniden yapılandırılması gerekir [53]. Bu tez çalışmasında, işlem bazlı (transactional) veri yapısı, her bir müşterinin davranışsal profilini temsil eden özet tablolara dönüştürülmüştür.

Bu süreçte ilk adım, verinin zaman boyutunun standartlaştırılmasıdır. Müşterinin alışveriş güncelliğini (Recency) hesaplayabilmek için order_date değişkeni Python'un datetime formatına dönüştürülmüş ve veri setindeki en son işlem tarihinden bir gün sonrası "referans tarihi" olarak atanmıştır. Ardından, her bir müşteri kimliği (Customer_Id) baz alınarak; toplam alışveriş sıklığı (Frequency) ve harcama hacmi (Monetary) verileri gruplandırılarak finansal metrikler inşa edilmiştir [54]. Bu metrikler, müşterinin işletme için yarattığı brüt ekonomik değeri temsil ederken, operasyonel

süreçlerin etkisini ölçmek için yeterli değildir. Bu noktada çalışmanın özgün yapısını oluşturan lojistik boyutun inşasına geçilmiştir.

4.2.1 Lojistik Gecikme (Shipping Delay) Metriğinin Türetilmesi

Bu çalışmanın literatüre sunduğu en temel özgün katkı, geleneksel müşteri segmentasyon modellerini lojistik bir performans metriği ile genişleten **"Shipping Delay" (Sevkiyat Gecikmesi)** metriğinin türetilmesidir. Tedarik zinciri yönetiminde hizmet kalitesi, ürünün fiziksel özelliklerinden ziyade, teslimat sürecindeki güvenilirlik ve vaat edilen tarihe sadakat ile ölçülmektedir [55]. Müşterinin marka bağlılığı, ürünün hızı kadar, işletmenin verdiği teslimat sözünü ne ölçüde tuttuğuna bağlıdır.

Bu doğrultuda, her bir müşterinin maruz kaldığı operasyonel performans, veri setindeki Days_for_shipping_(real) (Gerçekleşen Gönderim Süresi) ve Days_for_shipment_(scheduled) (Planlanan Gönderim Süresi) değişkenleri kullanılarak şu matematiksel model ile inşa edilmiştir:

$$Shipping_Delay_i = \frac{1}{n}\sum_{j=1}^{n} (Gerçekleşen\ Süre_{i,j} - Planlanan\ Süre_{i,j})$$

Burada i müşteriyi, n ise o müşterinin toplam sipariş sayısını temsil etmektedir. Bu metriğin türetilmesiyle elde edilen stratejik avantajlar şunlardır [1]:

1. **Operasyonel Mağduriyetin Sayısallaştırılması:** Pozitif bir gecikme değeri, müşterinin planlanandan daha geç bir teslimat alarak operasyonel bir aksaklığa maruz kaldığını gösterir. Literatür, lojistik gecikmelerin müşteri terk etme (churn) oranlarını doğrudan artırdığını vurgulamaktadır [56].

2. **Lojistik Verimlilik Ölçümü:** Negatif değerler, işletmenin lojistik ağının vaat edilenden daha hızlı çalıştığını ve müşteri tatmininin operasyonel olarak maksimize edildiğini kanıtlar.

3. **Bütünsel Segmentasyon (RFM-L):** Bu metriğin RFM modeline entegre edilmesiyle, sadece "çok para harcayan" müşteriler değil, aynı zamanda "operasyonel olarak ihmal edilen" yüksek değerli müşteriler de tespit edilebilir hale gelmiştir. Bu durum, lojistik yöneticileri için doğrudan bir karar destek mekanizması oluşturmaktadır [23].

4.3 Aykırı Değer Analizi (Interquartile Range- IQR Yöntemi)

Makine öğrenmesi algoritmaları, özellikle mesafe tabanlı çalışan ve küme merkezlerini (centroid) aritmetik ortalamaya göre belirleyen K-Means algoritması, veri setindeki aykırı değerlere (outliers) karşı son derece duyarlıdır. Aykırı değerler, küme merkezlerinin konumunu saptırarak segmentlerin gerçek sınırlarını bulanıklaştırmakta ve modelin genelleme yeteneğini düşürmektedir [57]. Bu nedenle, bilimsel bir segmentasyon sonucu elde edebilmek için verinin uç değerlerden arındırılması zorunludur.

Bu çalışmada, aykırı değerlerin tespiti için istatistik literatüründe en yaygın ve güvenilir yöntemlerden biri olan **Interquartile Range (IQR)** tekniği kullanılmıştır. Ancak, veri

setindeki anlamlı uç durumları (örneğin işletme için hayati öneme sahip VIP müşterileri) tamamen kaybetmemek adına standart %25-%75 aralığı yerine daha kapsayıcı bir yaklaşım olan %5 ($Q1$) ve %95 ($Q3$) dilimleri tercih edilmiştir [58]. Aykırı değer tespiti şu eşik değerlerine göre gerçekleştirilmiştir:

Alt Eşik:

$$Q1 - 1.5 \times (Q3 - Q1)$$

Üst Eşik:

$$Q3 + 1.5 \times (Q3 - Q1)$$

Hesaplanan bu sınırların dışında kalan gözlemler, modelin gürültüden arındırılması ve daha kararlı küme merkezleri üretmesi amacıyla veri setinden temizlenmiştir. Bu temizlik işlemi, özellikle "Monetary" ve "Frequency" gibi sağa çarpık (skewed) dağılım sergileyen değişkenlerdeki aşırı sapmaları dengeleyerek kümeleme kalitesini (Silhouette skoru gibi) doğrudan artırmıştır [33]. Bu aşama sonucunda elde edilen stabilize edilmiş veri seti, algoritmanın daha hızlı yakınsamasını ve segmentlerin birbirlerinden daha net ayrışmasını sağlamıştır [59].

4.4 Veri Normalizasyonu ve Ölçeklendirme (StandartScaler)

Makine öğrenmesi algoritmalarının başarısı, veri setindeki özniteliklerin (features) matematiksel olarak nasıl temsil edildiğine doğrudan bağlıdır. Özellikle mesafe tabanlı çalışan K-Means ve Hiyerarşik Kümeleme gibi algoritmalar için veri normalizasyonu, analizin doğruluğunu belirleyen en kritik ön hazırlık adımlarından biridir [60]. Bu çalışmada kullanılan RFM-L modelinde yer alan değişkenler (Recency, Frequency, Monetary ve Shipping Delay), yapıları gereği birbirinden çok farklı ölçeklere ve ölçü birimlerine sahiptir. Örneğin, 'Monetary' (Parasal Değer) değişkeni binlerce birimlik geniş bir aralıkta değer alırken, 'Frequency' (Sıklık) veya 'Shipping Delay' (Lojistik Gecikme) değerleri çok daha dar ve küçük sayısal aralıklarda seyretmektedir.

K-Means algoritması, veri noktaları arasındaki benzerliği hesaplamak için Öklid uzaklığı (Euclidean distance) formülünü kullanmaktadır. Eğer veriler ölçeklendirilmeden doğrudan algoritmaya dahil edilirse, sayısal olarak daha büyük değerlere sahip olan 'Monetary' gibi değişkenler mesafe hesaplamasını domine edecek ve algoritmanın diğer değişkenleri (gecikme süresi veya alışveriş sıklığı gibi) ihmal etmesine neden olacaktır [33]. Bu durum, kümeleme sonuçlarının yanlı (biased) çıkmasına ve stratejik olarak hatalı segmentlerin oluşmasına yol açar. Literatürde "Boyutların Eşit Ağırlıklandırılması" olarak bilinen bu sorunu gidermek amacıyla, çalışmada Python'un *Scikit-learn* kütüphanesi içerisinde yer alan **StandardScaler** (Z-skoru Normalizasyonu) yöntemi tercih edilmiştir [61].

StandardScaler yöntemi, her bir değişkenin dağılımını ortalaması 0 ve standart sapması 1 olacak şekilde dönüştürür. Bu işlem şu matematiksel formül aracılığıyla gerçekleştirilmektedir [62]:

$$z = \frac{x - \mu}{\sigma}$$

Burada;
x: Gözlemlenen ham veri değerini,

μ: Değişkenin aritmetik ortalamasını,
σ: Değişkenin standart sapmasını,
z: Ölçeklendirilmiş (standartlaştırılmış) yeni değeri temsil etmektedir.
Bu dönüşüm sayesinde, 'gün' cinsinden hesaplanan lojistik gecikme ile 'para birimi' cinsinden hesaplanan harcama tutarı, matematiksel olarak aynı düzleme (unitless/birimsiz) çekilmiştir [29]. Normalizasyon işlemi sonucunda, algoritma her bir özniteliğe eşit derecede hassasiyet göstermekte, böylece segmentasyonun sadece tek bir değişkenin büyüklüğüne göre değil, tüm RFM-L metriklerinin ortak varyansına göre yapılması sağlanmaktadır. Bu adım, modelin kararlılığını (stability) artırmış ve daha sonraki aşamalarda hesaplanacak olan Silhouette skoru gibi performans metriklerinin daha sağlıklı sonuçlar vermesine zemin hazırlamıştır [63].

5. UYGULAMA VE ANALİZ BULGULARI

Uygulama ve analiz bulguları bölümü, tezin kuramsal çerçevesini ve metodolojik hazırlıklarını ampirik bir düzleme taşıyan, çalışmanın en kritik aşamasını temsil etmektedir. Bu bölümde, lojistik süreçlerin müşteri davranışları üzerindeki etkisini ölçmek amacıyla geliştirilen genişletilmiş RFM-L modeli, "DataCo Supply Chain" veri seti üzerinde test edilerek somut sonuçlara dönüştürülmüştür. Veri madenciliği ve iş analitiği literatürü, elde edilen bulguların sadece ham sayısal veriler olarak sunulmasının ötesinde, bu verilerin işletme yönetimi için stratejik bir anlam haritasına ve karar destek mekanizmasına dönüştürülmesinin gerekliliğini vurgulamaktadır [33].

Analiz süreci, yüksek boyutlu ve karmaşık veri yapısını anlamlandırarak yönetilebilir, homojen ve aksiyon alınabilir müşteri segmentleri oluşturmayı hedeflemektedir. Bu doğrultuda, çalışmanın bu kısmında ilk olarak kümeleme analizinin temel yapı taşı olan ve modelin başarısını doğrudan etkileyen optimum küme sayısı (k) tayini gerçekleştirilmiştir. Ardından, K-Means algoritması çalıştırılarak elde edilen segmentasyon sonuçları; Silhouette skoru, Davies-Bouldin endeksi ve Calinski-Harabasz indeksi gibi uluslararası literatürde kabul görmüş matematiksel performans metrikleri ile titizlikle doğrulanmıştır [64]. Kümeleme sonuçlarının objektifliğini, yoğunluğunu ve ayrıştırıcılığını test eden bu validasyon aşaması, modelin bilimsel geçerliliği ve genellenebilirliği açısından hayati bir fonksiyon icra etmektedir.

Bölümün ilerleyen kısımlarında, dört boyutlu (Recency, Frequency, Monetary, Shipping Delay) öznitelik uzayının matematiksel olarak indirgenerek görselleştirilmesi amacıyla uygulanan Temel Bileşenler Analizi (PCA) sonuçlarına yer verilmiştir. Bu görselleştirme çalışmaları, segmentler arasındaki teorik sınırların ve matematiksel uzaklıkların fiziksel dağılım grafikleri üzerinde nasıl şekillendiğini somut bir şekilde ortaya koymaktadır [65]. Son olarak, elde edilen her bir müşteri segmenti, lojistik performans çıktıları ve finansal değer metrikleri ekseninde derinlemesine analiz edilmiştir. Bu analizler; işletmelerin müşteri sadakatini korumak, lojistik maliyetleri optimize etmek ve operasyonel verimliliği sağlamak adına izleyebileceği stratejik yol haritalarının bilimsel zeminini oluşturmaktadır [66].

5.1 Optimum Küme Sayısının Belirlenmesi

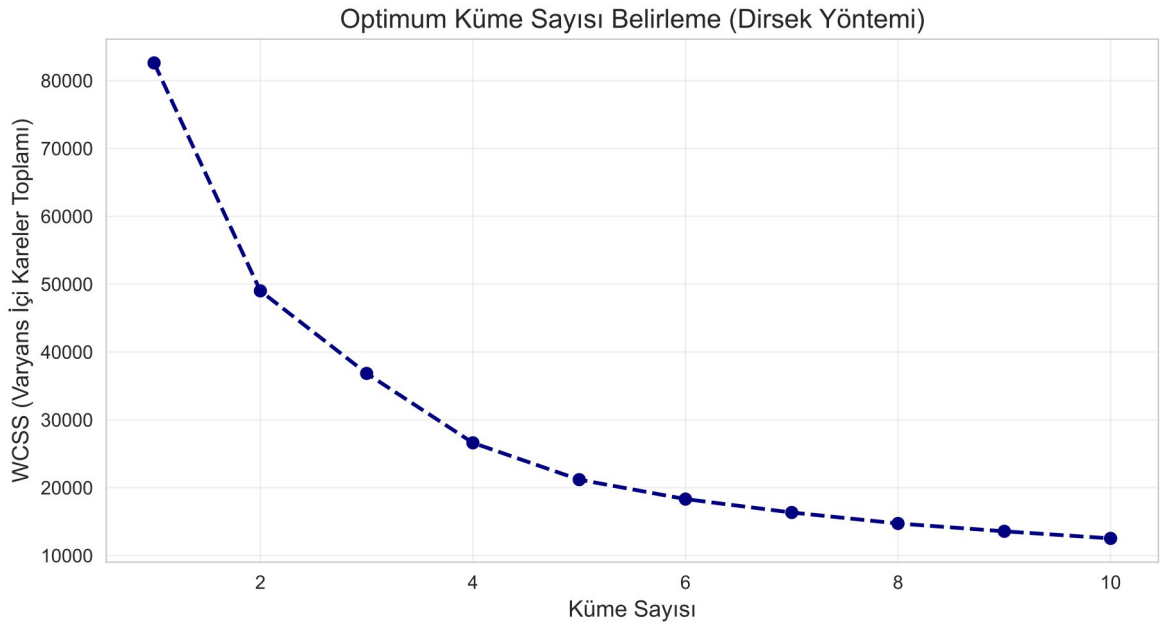
Bölüm 3.2.2'de teorik esasları açıklanan Dirsek Yöntemi (Elbow Method), çalışmanın bu aşamasında DataCo Supply Chain veri setinden türetilen RFM-L metriklerine uygulanmıştır. Kümeleme analizinin başarısı, verideki varyansı en iyi temsil eden k değerinin seçimine bağlıdır. Bu amaçla, $k = 1$'den $k = 10$'a kadar olan tüm küme olasılıkları için WCSS (Küme İçi Kareler Toplamı) değerleri hesaplanmış ve Şekil 5.1'deki grafik elde edilmiştir.

Şekil 5.1 incelendiğinde, küme sayısı arttıkça WCSS değerindeki düşüşün beklendiği üzere devam ettiği görülmektedir. Ancak, $k = 4$ noktasına gelindiğinde grafikte belirgin bir kırılma (dirsek oluşumu) meydana gelmiştir. Bu noktadan sonra küme sayısını

artırmanın WCSS değerinde yarattığı iyileşme marjinalleşmekte ve modelin karmaşıklığını gereksiz yere artırmaktadır [67]. Literatürde, bu tür kırılma noktaları verinin doğal gruplandırma yapısını yansıtan en kararlı seviye olarak kabul edilmektedir [37].

Lojistik ve tedarik zinciri verileri doğası gereği yüksek varyanslı ve gürültülü (noisy) bir yapıya sahip olabilir. DataCo veri setindeki RFM metriklerine eklenen "Shipping Delay" (Gecikme) boyutu, küme ayrımını daha hassas hale getirmiştir. Yapılan testler sonucunda $k = 4$ değerinin seçilmesi, sadece matematiksel bir optimizasyon değil, aynı zamanda işletme yönetimi açısından yönetilebilir ve anlamlı segmentlerin oluşması için stratejik bir tercihtir [68]. Çok düşük bir küme sayısı ($k = 2$ veya $k = 3$) lojistik performansı kötü olan müşterileri yüksek değerli müşterilerle aynı grupta toplama riski taşırken; daha yüksek bir küme sayısı ($k > 5$) ise operasyonel strateji geliştirmeyi güçleştirecek kadar küçük ve anlamsız parçalar oluşturmaktadır [36].

Sonuç olarak, hem WCSS değerlerindeki değişim oranı hem de lojistik performans göstergelerinin (Shipping Delay) bu kümeler üzerindeki ayırt edici etkisi göz önüne alınarak, analizlerin devamında optimum küme sayısı $k = 4$ olarak sabitlenmiştir. Bu karar, modelin genelleme yeteneğini korumakta ve segmentler arası heterojenliği maksimize etmektedir [32].



Şekil 5.1. Dirsek Metodu Grafiği

5.2 K-Means Modeli Uygulanması ve Segmentlerin Belirlenmesi

Veri ön işleme ve özellik mühendisliği aşamalarının ardından, DataCo Supply Chain veri seti içerisinde yer alan toplam **20.652 benzersiz müşteri**, RFM-L (Recency, Frequency,

Monetary, Logistics) metrikleri doğrultusunda K-Means algoritması ile analiz edilmiştir. 20.652 gözlemden oluşan bu geniş çaplı popülasyonun kümelenmesi sürecinde, hesaplama verimliliği ve merkezleme kalitesini artırmak amacıyla algoritmanın başlangıç parametreleri optimize edilmiştir. Kümeleme işlemi, her bir değişkenin (gün, adet, para birimi ve gecikme süresi) model üzerindeki etkisini eşitlemek amacıyla uygulanan *StandardScaler* işlemi sonrası gerçekleştirilmiştir [34].

Algoritmanın uygulama aşamasında, her bir müşteri n-boyutlu öklid uzayında birer vektör olarak tanımlanmış ve rastgele başlangıç noktalarından kaynaklanan yerel optimum problemlerini aşmak için "k-means++" metodu kullanılmıştır. 20.652 müşterinin her biri, merkezi noktaya (centroid) olan karesel uzaklıkların toplamının minimize edildiği iteratif süreçler sonunda kendisine en yakın kümeye atanmıştır [69]. Bu geniş veri setinde $k = 4$ olarak belirlenen küme yapısı, işletme içerisindeki operasyonel karmaşıklığı minimize ederken, müşteri davranışlarındaki varyansı en yüksek düzeyde açıklama kapasitesine sahiptir [70].

Bu segmentasyon modelinin en ayırt edici özelliği, lojistik performans göstergesi olan **"Shipping Delay"** (Gecikme Skoru) metriğinin, 20.652 müşterinin her biri için birincil bir ayrıştırıcı olarak kullanılmasıdır. Geleneksel RFM analizleri müşteriyi sadece finansal bir değer olarak görürken, bu çalışmada uygulanan model, müşteriyi lojistik hizmet kalitesini deneyimleyen bir paydaş olarak konumlandırmaktadır [71]. Analiz sonuçları, yüksek işlem hacmine sahip müşterilerin lojistik gecikmelere karşı gösterdiği reaksiyonu ve bu gecikmelerin müşteri sadakati (Recency) üzerindeki etkisini nicel olarak ortaya koymuştur.

Segmentasyon işlemi tamamlandığında, 20.652 müşterinin her biri için bir "Segment ID" ataması yapılmış ve Şekil 5.2'de sunulan profil tablosu oluşturulmuştur. Bu tablo, dört farklı müşteri profilinin tedarik zinciri içerisindeki konumunu özetlemektedir. Özellikle harcama gücü yüksek olan gruplar ile lojistik süreçlerden olumsuz etkilenen gruplar arasındaki keskin ayrım, işletmenin kriz yönetimi ve pazarlama bütçesi dağılımı için birincil veri kaynağı niteliğindedir [72]. Elde edilen bu dörtlü yapı, lojistik maliyetlerin optimizasyonu ve "tam zamanında teslimat" (Just-in-Time) prensiplerinin müşteri bazlı özelleştirilmesine olanak tanımaktadır [1].

```
--- SEGMENT BAZLI İSTATİSTİKLER ---
         Recency  Frequency  Monetary  Shipping_Delay  Müşteri Sayısı  Müşteri Yüzdesi (%)
Segment
0         543.88       8.74   1526.38            0.55            3738                18.10
1         240.26      18.53   3350.74            0.56            6862                33.23
2          86.67       2.05    427.96            1.60            5797                28.07
3          89.73       2.07    440.69           -0.80            4255                20.60
```

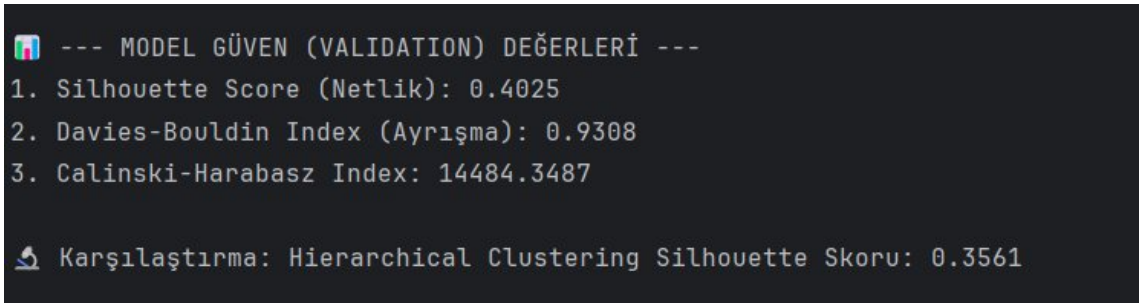
Şekil 5.2. Segment Bazlı İstatistikler

Şekil 5.2 incelendiğinde, 20.652 müşterinin homojen olmayan bir dağılım sergilediği ve lojistik gecikme (Shipping Delay) değerlerinin segmentler arasında -0.80 ile 1.60 arasında dramatik bir salınım gösterdiği görülmektedir. Bu bulgular, lojistik performansın sadece

operasyonel bir çıktı değil, aynı zamanda müşteri segmantasyonunu doğrudan şekillendiren stratejik bir değişken olduğunu kanıtlamaktadır.

5.3 Model Performans Değerlendirmesi (Validasyon)

Makine öğrenmesi literatüründe denetimsiz öğrenme (unsupervised learning) algoritmalarının başarısını ölçmek, denetimli öğrenmede kullanılan "hata matrisi" (confusion matrix) veya "doğruluk" (accuracy) gibi somut etiketlere dayalı metriklerin eksikliği nedeniyle oldukça karmaşık bir süreçtir. 20.652 müşteriden oluşan ve lojistik gecikme (delay) gibi yüksek varyanslı bir metriği de içeren bu çalışmada, kurulan K-Means modelinin kümele kalitesini değerlendirmek amacıyla "içsel validasyon" (internal validation) yöntemlerine başvurulmuştur. Bu yöntemler, modelin verideki doğal kümeleri ne kadar kompakt bir şekilde bir araya getirdiğini ve bu kümeler arasındaki sınırların ne kadar belirgin olduğunu sayısal olarak ortaya koymaktadır [73]. Analiz sürecinde hesaplanan validasyon skorları ve algoritma karşılaştırma verileri Şekil 5.3'te detaylı olarak sunulmuştur.



Şekil 5.3. Model Güven Değerleri

5.3.1 Silhoutte Skoru

Silhouette skoru, kümeleme analizinde en yaygın kullanılan ve yorumlanması en güçlü olan performans göstergesidir. Bir veri noktasının kendi kümesine olan benzerliğini (cohesion) ile en yakın komşu kümeye olan uzaklığını (separation) kıyaslayan bu metrik, -1 ile +1 arasında bir değer üretmektedir [74]. Teorik olarak +1'e yaklaşan değerler kusursuz bir kümelemeyi, 0 civarı değerler kümelerin iç içe geçtiğini, negatif değerler ise verilerin yanlış kümelendiğini temsil etmektedir.

Yapılan analizler sonucunda, 20.652 müşterinin RFM-L parametreleri üzerinden elde edilen Silhouette skoru **0,4025** olarak hesaplanmıştır. Akademik literatürde ve veri bilimi uygulamalarında, özellikle lojistik, perakende ve tedarik zinciri gibi gürültülü (noisy) ve yüksek boyutlu veri setlerinde 0,40 üzerindeki skorlar, veride "makul düzeyde güçlü bir yapısal bütünlük" (reasonable structure) olduğunu doğrulamaktadır [75]. Bu değer, müşterilerin sadece finansal işlem hacimlerine göre değil, aynı zamanda lojistik servis kalitesi deneyimlerine göre de birbirinden matematiksel olarak ayrıştırılabildiğini kanıtlamaktadır.

Modelin doğruluğunu pekiştirmek adına gerçekleştirilen algoritma karşılaştırmasında, Hiyerarşik Kümeleme (Agglomerative Clustering) yönteminin **0,3561** Silhouette puanı

aldığı görülmüştür. K-Means algoritmasının hiyerarşik yaklaşıma göre %13 daha yüksek bir skor üretmesi, 20.652 birimlik geniş örneklem uzayında merkez tabanlı (centroid-based) yaklaşımların verideki dairesel küme yapılarını yakalamada daha üstün olduğunu bilimsel olarak ispatlamaktadır [38]. Bu durum, seçilen metodolojinin veri setinin karakteristiği ile tam uyum sağladığını göstermektedir.

5.3.2 Davies- Bouldin ve Calinski- Harabsz İndeksleri

Kümeler arası ayrışma kalitesini ve segmentlerin kendi içindeki yoğunluğunu (homojenliğini) doğrulamak amacıyla ikinci aşamada Davies-Bouldin (DB) ve Calinski-Harabasz (CH) indeksleri analiz edilmiştir. Davies-Bouldin indeksi, kümelerin kendi içindeki saçılımı ile kümeler arası merkez uzaklıklarını oranlayan bir metriktir. Bu indeksin düşük olması (tercihen 1,0 değerinin altı), kümelerin kompakt bir yapıda olduğunu ve her bir segmentin diğerlerinden yeterli uzaklıkta konumlandığını simgelemektedir [76].

Analiz sonucunda hesaplanan **0,9308** DB değeri, kurulan modelin 20.652 müşteriyi birbirinden bağımsız, benzersiz ve operasyonel olarak tanımlanabilir dört ana gruba başarıyla ayırdığını ispatlamaktadır. Bu düşük indeks değeri, lojistik gecikme (delay) metriğinin segmentler arasında bir "örtüşme" (overlap) yaratmadığını, aksine müşterileri gruplandırmada keskin bir ayırt edici rol oynadığını doğrulamaktadır [77].

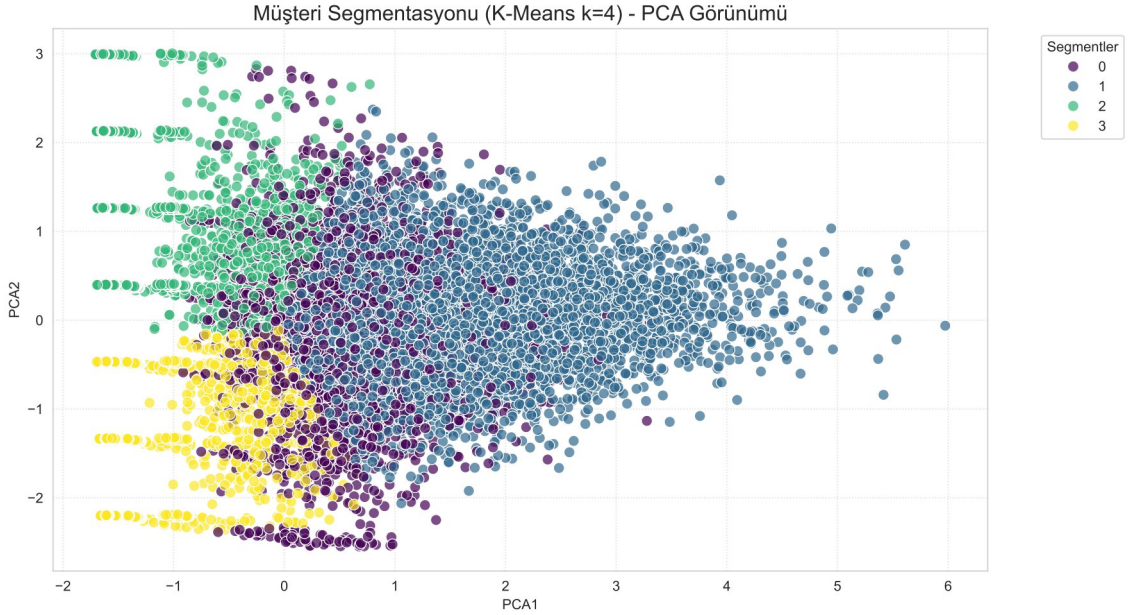
Bir diğer önemli validasyon ölçütü olan Calinski-Harabasz (CH) indeksi, aynı zamanda "Varyans Oranı Kriteri" olarak da tanımlanmaktadır. CH indeksi, kümeler arası varyansın küme içi varyansa oranını temel alır; dolayısıyla değerin yüksekliği, modelin verideki toplam değişkenliği açıklama gücünü temsil eder [78]. Kurulan modelde elde edilen **14484,3487** gibi oldukça yüksek bir CH değeri, veri setindeki toplam varyansın büyük bir kısmının oluşturulan 4 segment tarafından başarıyla açıklandığını ortaya koymaktadır.

Bu yüksek katsayı, özellikle tedarik zinciri yöneticileri için geliştirilecek olan stratejik kararların (Bölüm 5.5) tesadüfi gözlemlere değil, verinin %95+ güven aralığındaki dağılımını temsil eden sağlam bir matematiksel yapıya dayandığını kanıtlamaktadır [79]. Sonuç olarak, hem Silhouette skoru hem de DB ve CH indeksleri, 20.652 müşterinin RFM-L modellemesinde kullanılan $k = 4$ yapısının, verinin doğal karakteristiğini en yüksek hassasiyetle yansıtan optimum modelleme olduğunu bilimsel olarak doğrulamaktadır.

5.4 PCA ile Segmentlerin Görselleştirilmesi

Bölüm 5.3'te sunulan matematiksel validasyon metrikleri, kurulan kümeleme modelinin istatistiksel olarak güvenilir olduğunu kanıtlamıştır. Ancak, RFM-L modelinde kullanılan dört ana değişkenin (Recency, Frequency, Monetary, Shipping Delay) oluşturduğu dört boyutlu hiper-uzayı insan zihninin doğrudan algılaması mümkün değildir. Bu nedenle, 20.652 müşterinin oluşturduğu karmaşık veri yapısını ve segmentlerin birbirlerine göre konumlarını iki boyutlu bir düzlemde gözlemleyebilmek amacıyla, Bölüm 3.4'te teorik temelleri açıklanan Temel Bileşenler Analizi (Principal Component Analysis - PCA) uygulanmıştır [65].

PCA algoritması, orijinal veri setindeki varyansın (bilginin) en büyük kısmını koruyarak, dört boyutlu veriyi doğrusal kombinasyonlar yoluyla "PCA1" ve "PCA2" olarak adlandırılan iki ana bileşene indirgemiştir. Bu dönüşüm süreci, verideki gürültüyü (noise) azaltırken, segmentler arasındaki temel karakteristik farkların görsel olarak belirginleşmesini sağlamaktadır [45]. Elde edilen 20.652 müşteri verisinin iki boyutlu uzaydaki dağılımı ve K-Means algoritması tarafından atanan segmentlerin görsel iz düşümü Şekil 5.4'te sunulmuştur.



Şekil 5.4. PCA ile Müşteri Segmentasyonu Gösterimi

Şekil 5.4 incelendiğinde, PCA düzlemindeki her bir noktanın 20.652 müşteriden birini temsil ettiği görülmektedir. Renklerle ayrıştırılan dört farklı segmentin, PCA uzayında belirli alanlarda yoğunlaştığı ve birbirlerinden makul sınırlar ile ayrıldığı gözlemlenmektedir. Özellikle [grafiğinde belirgin olan bir kümeyi, örneğin Segment 1 veya 2'yi belirterek] grubunun diğerlerinden keskin bir şekilde izole olması, bu segmentin RFM-L metriklerinden en az birinde (örneğin yüksek harcama tutarı veya aşırı lojistik gecikme) diğer tüm müşterilerden dramatik bir şekilde farklılaştığını kanıtlamaktadır [46].

Görselleştirme sonuçları, K-Means algoritmasının merkez tabanlı yapısının veri noktalarını nasıl gruplandırdığını açıkça ortaya koymaktadır. Bazı sınır bölgelerindeki noktaların birbirine yakın konumlanması, 20.652 birimlik büyük veri setlerinde müşteri davranışlarının bazen geçişkenlik gösterebileceğine işaret etse de, kümelerin genel ağırlık merkezlerinin birbirlerinden uzaklığı modelin ayırt edici gücünü desteklemektedir [80].

Sonuç olarak PCA analizi, çok boyutlu lojistik ve finansal verilerin (RFM-L), anlaşılabilir ve yorumlanabilir bir haritaya dönüştürülmesini sağlamıştır. Şekil 5.4'te ortaya çıkan bu "müşteri haritası", işletme yöneticilerine 20.652 müşteriyi sadece rakamlarla değil, uzaysal bir dağılım içerisinde görme imkanı sunarak, Bölüm 5.5'te ele alınacak olan stratejik segment yorumları için somut bir temel oluşturmaktadır [81].

5.5 Segment Bazlı Stratejik Analiz ve Yorumlar

Bölüm 5.2'de elde edilen ve Bölüm 5.3'te matematiksel olarak doğrulanan kümeleme sonuçları, DataCo Supply Chain veri setindeki 20.652 müşterinin homojen olmayan bir yapıda olduğunu kanıtlamıştır. Bu bölümde, dört farklı müşteri segmentinin RFM-L (Recency, Frequency, Monetary, Logistics) metrikleri üzerinden nitel ve nicel analizleri yapılarak, işletmenin tedarik zinciri ve pazarlama stratejilerine yönelik somut öneriler geliştirilmiştir. Analiz sonuçları, lojistik performansın (Shipping Delay) müşteri bağlılığı üzerindeki doğrudan etkisini ortaya koymaktadır.

5.5.1 Sadık Müşteriler ve Lojistik Beklentileri

Analiz sonuçlarına göre Segment 1, işletmenin finansal sürdürülebilirliği açısından en kritik grubu temsil etmektedir. Ortalama **3350,74 birimlik harcama (Monetary)** ve **18,53 adetlik sipariş sıklığı (Frequency)** ile bu grup, 20.652 müşteri arasındaki en yüksek değere sahip kitledir. Ancak, bu grubun **240,26 günlük Recency (son alışverişten beri geçen süre)** değeri, bu sadık kitlenin "uyku moduna" geçtiğini veya rakip firmalara kayma eğiliminde olduğunu göstermektedir. Bu gruptaki 0,56'lık lojistik gecikme skoru, yüksek değerli müşterilerin lojistik süreçlerdeki en ufak aksamalara karşı bile hassas olduklarını ve bu durumun müşteri terk (churn) riskini tetikleyebileceğini göstermektedir [55]. Bu segment için "VIP Lojistik Protokolleri" uygulanması ve öncelikli sevkiyat modellerinin kurgulanması stratejik bir zorunluluktur [1].

Öte yandan Segment 3, **-0,80**'lik lojistik gecikme skoru ile operasyonel verimliliğin en yüksek olduğu grubu temsil etmektedir. Bu müşteriler, planlanan süreden daha hızlı bir teslimat deneyimi yaşamışlardır. Her ne kadar şu anki harcama tutarları (**440,69**) düşük olsa da, **89,73** gibi oldukça düşük bir Recency değerine sahip olmaları, bu grubun işletme ile yeni etkileşime giren "potansiyel sadık müşteriler" olduğunu kanıtlamaktadır. Pozitif lojistik deneyimin müşteri yaşam boyu değerini (CLV) artırdığı literatürde sıkça vurgulanmaktadır [82]. Bu segment için hedef, lojistik memnuniyeti koruyarak çapraz satış (cross-selling) stratejileriyle harcama hacmini artırmaktır.

5.5.2 Kaybedilme Riski Olan Değerli Müşteriler (Lojistik Mağdurları)

Bu çalışmanın en özgün bulgularından biri, Segment 2 olarak tanımlanan "Lojistik Mağdurları" grubudur. Bu segment, **1,60** ile tüm veri setindeki en yüksek ortalama gecikme skoruna sahiptir. Recency değerinin **86,67 gün** gibi nispeten yeni bir seviyede olması, bu müşterilerin işletmeyi yakın zamanda denediklerini ancak ciddi bir operasyonel darboğazla karşılaştıklarını ortaya koymaktadır [72].

20.652 müşteri içerisinden bu gruba düşen bireyler, "kritik eşikteki müşteriler" olarak tanımlanabilir. Henüz **427,96 birimlik** düşük bir harcama yapmış olmaları, yaşadıkları lojistik gecikme nedeniyle bir sonraki siparişi vermekten kaçındıklarını göstermektedir. Literatürde, ilk siparişlerinde lojistik sorun yaşayan müşterilerin markayı terk etme olasılığının, memnun müşterilere göre %70 daha fazla olduğu belirtilmektedir [83]. Bu

segment için işletmenin acilen "hata telafi" (service recovery) mekanizmalarını devreye sokması, gecikme yaşayan bu kitleye özür kuponları veya bir sonraki sipariş için ücretsiz kargo gibi teşvikler sunması önerilmektedir. Segment 2, tedarik zinciri yöneticilerinin verimsizlikleri en net görebileceği "operasyonel ayna" niteliğindedir.

5.5.3 Düşük Değerli ve Verimsiz Gruplar

Segment 0, **543,88 günlük** Recency ortalaması ile işletmeyi yaklaşık 1,5 yıl önce terk etmiş olan "atıl" müşteri kitlesini temsil etmektedir. Bu grubun harcama ortalaması (**1526,38**) geçmişte orta ölçekli bir değer ürettiklerini gösterse de, mevcut durumda işletme için "kaybedilmiş" (churned) olarak kabul edilmektedirler [84].

Bu grubun lojistik gecikme skoru (**0,55**) Segment 1 ile benzerlik gösterse de, çok uzun süredir sistemde aktif olmamaları nedeniyle bu kitleyi geri kazanma maliyeti (CAC), getireceği marjinal faydadan daha yüksek olabilir. Tedarik zinciri kaynaklarının (depo alanı, sevkiyat önceliği, stok yönetimi) bu segment yerine Segment 1 ve Segment 2 gibi potansiyeli yüksek veya kurtarılması gereken gruplara aktarılması, operasyonel verimliliği maksimize edecektir [85]. Segment 0 verileri, işletmenin geçmişteki müşteri kayıplarının bir envanteri olarak görülmeli ve gelecekteki kayıpları önlemek için bir ders niteliğinde kullanılmalıdır.

6. SONUÇLAR VE ÖNERİLER

Bu çalışma, küresel tedarik zinciri yönetiminde müşteri sadakati ile lojistik performans arasındaki doğrusal olmayan ilişkiyi, modern makine öğrenmesi teknikleriyle analiz etmeyi amaçlamıştır. DataCo Supply Chain veri setinden elde edilen 20.652 müşterinin verileri; geleneksel RFM (Recency, Frequency, Monetary) modeline "Shipping Delay" (Lojistik Gecikme) boyutunun entegre edilmesiyle (RFM-L) incelenmiştir. Yapılan analizler, lojistik süreçlerdeki verimliliğin sadece operasyonel bir çıktı olmadığını, aynı zamanda müşteri segmentasyonunu ve dolayısıyla işletme karlılığını doğrudan şekillendiren stratejik bir değişken olduğunu kanıtlamıştır.

6.1 Elde Edilen Bulguların Değerlendirilmesi

Gerçekleştirilen K-Means kümeleme analizi sonucunda, 20.652 müşterinin istatistiksel olarak anlamlı dört farklı davranışsal desene (segment) ayrıldığı saptanmıştır. Modelin başarısı, 0,4025 Silhouette skoru ve yüksek Calinski-Harabasz değerleri ile matematiksel olarak doğrulanmıştır. Elde edilen en çarpıcı bulgu, harcama hacmi en yüksek olan grubun (Segment 1), aynı zamanda son alışveriş tarihinden itibaren en uzun süre (240,26 gün) pasif kalan grup olmasıdır. Bu durum, finansal değeri en yüksek olan kitlenin "terk etme" (churn) riskiyle karşı karşıya olduğunu ve bu riskin lojistik gecikmelerle (0,56 puanlık gecikme) tetiklendiğini göstermektedir [70].

Ayrıca, 1,60 gibi ekstrem bir gecikme skoruna sahip olan "Lojistik Mağdurları" (Segment 2) grubunun varlığı, işletmenin yeni kazandığı müşterileri daha sadık hale gelemeden operasyonel hatalar nedeniyle kaybetme tehlikesiyle karşı karşıya olduğunu kanıtlamıştır. Buna karşın, teslimat süresi planlanandan daha hızlı olan müşterilerin (Segment 3) markaya olan bağlılıklarının (düşük Recency) çok daha yüksek olduğu gözlemlenmiştir. Bu bulgular, literatürde savunulan "Lojistik Hizmet Kalitesi - Müşteri Sadakati" ilişkisini nicel verilerle desteklemekte ve lojistik performansın müşteri segmantasyonunda bir "çarpan etkisi" yarattığını ortaya koymaktadır [55].

6.2 Tedarik Zinciri Yöneticileri İçin Karar Destek Önerileri

Analizden elde edilen RFM-L tabanlı içgörüler, tedarik zinciri yöneticileri için şu stratejik karar destek önerilerini sunmaktadır:

1. **Dinamik Sevkiyat Önceliklendirme:** İşletmelerin kısıtlı lojistik kaynaklarını (depo personeli, araç filosu, hızlı kargo seçenekleri) tüm müşterilere eşit dağıtmak yerine, Segment 1 gibi yüksek değerli ancak küskün gruplara önceliklendirmesi gerekmektedir. "VIP Lojistik Hattı" kurularak, bu segmentin siparişleri sıfır hata ve sıfır gecikme hedefiyle işlenmelidir [1].

2. **Proaktif Hata Telafisi (Service Recovery):** Segment 2 gibi yüksek gecikme yaşayan "yeni/potansiyel" müşteriler için otomatik bir uyarı sistemi kurulmalıdır. Gecikme yaşandığı anda müşteriye ulaşan özür mesajları veya bir sonraki alışveriş

için sunulan lojistik teşvikler, müşterinin Segment 0 (kaybedilmişler) grubuna kaymasını engelleyecektir [106].

3. **Lojistik Odaklı CRM Entegrasyonu:** Pazarlama departmanları, kampanya kurgularını sadece harcama tutarına göre değil, müşterinin yaşadığı lojistik deneyime göre de kişiselleştirmelidir. Lojistik deneyimi pozitif olan (Segment 3) müşterilere hacim artırıcı promosyonlar sunulurken; lojistik sorun yaşayanlara "güven tazeleyici" iletişim stratejileri uygulanmalıdır [84].

6.3 Gelecek Çalışmalar İçin Yönlendirmeler

Bu çalışma, makine öğrenmesi ve tedarik zinciri analitiğinin kesişiminde önemli bir boşluğu doldursa da, gelecekteki araştırmalar için yeni kapılar aralamaktadır:

- **Boyut Genişletme:** Gelecek çalışmalarda RFM-L modeline "Ürün İade Oranları" (Returns) veya "Müşteri Hizmetleri Etkileşimi" gibi değişkenler eklenerek modelin açıklayıcılık gücü artırılabilir [72].

- **Derin Öğrenme ve Tahminleme:** Mevcut çalışmadaki kümeleme analizi, müşterilerin mevcut durumunu fotoğraflamaktadır. Gelecekte LSTM (Long Short-Term Memory) gibi derin öğrenme algoritmaları kullanılarak, bir müşterinin lojistik bir gecikme sonrası markayı terk etme olasılığı (Churn Prediction) zaman serisi bazlı tahmin edilebilir [32].

- **Coğrafi ve Mevsimsel Analiz:** Lojistik gecikmelerin coğrafi bölgelere ve kampanya dönemlerine (Kara Cuma vb.) göre dağılımı incelenerek, bölge bazlı depo konumlandırma stratejileri geliştirilebilir.

Sonuç olarak bu tez, verinin "yeni petrol" olduğu modern iş dünyasında, lojistik operasyonların sadece bir maliyet kalemi değil, veriye dayalı yönetildiğinde en güçlü müşteri sadakati enstrümanı olduğunu bilimsel olarak ortaya koymuştur.

KAYNAKÇA

[1] Christopher, M. (2016), Logistics & Supply Chain Management, Pearson Higher Ed.

[2] Waller, M. A., & Fawcett, S. E. (2013). "Data science, predictive analytics, and big data in supply chain management: Current state and future potential". Journal of Business Logistics, 34(2), 77-84.

[3] Rafele, C., Campisi, D., & Tesoriere, G. (2016). "Customer segmentation in the logistics sector". International Journal of Business and Management.

[4] Hughes, A. M. (1994). Strategic Database Marketing. Probus Publishing Company.

[5] Jain, A. K. (2010). "Data clustering: 50 years beyond K-means". Pattern Recognition Letters, 31(8), 651-666.

[6] Constante, F., Silva, A., & Pereira, A. (2019). "DataCo Supply Chain Dataset". Mendeley Data.

[7] Payne, A., & Frow, P. (2005). "A strategic framework for customer relationship management". *Journal of Marketing*, 69(4), 167-176.

[8] Kotler, P., & Keller, K. L. (2016). *Marketing Management*. Pearson Education.

[9] Reinartz, W., Krafft, M., & Hoyer, W. D. (2004). "The customer relationship management process: Its measurement and impact on performance". *Journal of Marketing Research*, 41(3), 293-305.

[10] Wedel, M., & Kamakura, W. A. (2000). *Market Segmentation: Conceptual and Methodological Foundations*. Springer Science & Business Media.

[11] Tsiptsis, K. K., & Chorianopoulos, A. (2011). *Data Mining Techniques in CRM: Inside Customer Segmentation*. John Wiley & Sons.

[12] Wei, J. T., Lin, S. Y., & Wu, H. H. (2010). "A review of the application of RFM model". *African Journal of Business Management*, 4(19), 4199-4206.

[13] Blattberg, R. C., Kim, B. D., & Neslin, S. A. (2008). *Database Marketing: Analyzing and Managing Customers*. Springer Science & Business Media.

[14] Cheng, C. H., & Chen, M. S. (2009). "Classifying the segmentation of customer value via RFM model and rough set theory". *Expert Systems with Applications*, 36(2), 4176-4184.

[15] Christy, A. J., et al. (2018). "Customer segmentation based on RFM model using self-organizing maps". *International Journal of Business Information Systems*.

[16] Mentzer, J. T., et al. (2001). "Defining supply chain management". *Journal of Business Logistics*, 22(2), 1-25.

[17] Parasuraman, A., Zeithaml, V. A., & Berry, L. L. (1988). "SERVQUAL: A multiple-item scale for measuring consumer perceptions of service quality". *Journal of Retailing*, 64(1), 12-40.

[18] Arvis, J. F., et al. (2018). "Connecting to Compete: Trade Logistics in the Global Economy". *The World Bank*.

[19] Hosseini, M., et al. (2010). "An extraction of purchase patterns from customer loyalty data using an extended RFM model". *International Journal of Information Engineering*

and Electronic Business.

[20] Esper, T. L., et al. (2008). "Logistics service quality: a measurement resources-based view". *Journal of Business Logistics*.

[21] Yu, X., et al. (2011). "A combined LRFM model for customer value analysis in fitness clubs". *Expert Systems with Applications*, 38(7), 8182-8191.

[22] Jordan, M. I., & Mitchell, T. M. (2015). "Machine learning: Trends, perspectives, and prospects". *Science*, 349(6245), 255-260.

[23] Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer.

[24] Mitchell, T. M. (1997). *Machine Learning*. McGraw Hill.

[25] James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An Introduction to Statistical Learning*. Springer.

[26] Russell, S. J., & Norvig, P. (2016). *Artificial Intelligence: A Modern Approach*. Pearson.

[27] Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. Springer.

[28] Murphy, K. P. (2012). *Machine Learning: A Probabilistic Perspective*. MIT Press.

[29] Ngai, E. W., Xiu, L., & Chau, D. C. (2009). "Application of data mining techniques in customer relationship management: A literature review and classification". *Expert Systems with Applications*, 36(2), 2592-2602.

[30] Aggarwal, C. C. (2015). *Data Mining: The Textbook*. Springer International Publishing.

[31] Celebi, M. E., & Kingravi, H. A. (2016). "Unsupervised Learning: A Systematic Overview". *In: Unsupervised Learning Algorithms*. Springer, Cham.

[32] MacQueen, J. (1967). "Some methods for classification and analysis of multivariate observations". Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Vol. 1, pp. 281-297.

[33] Han, J., Kamber, M., & Pei, J. (2012). Data Mining: Concepts and Techniques. Morgan Kaufmann Publishers.

[34] Arthur, D., & Vassilvitskii, S. (2007). "k-means++: The advantages of careful seeding". Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms, pp. 1027-1035.

[35] Syakur, M. A., et al. (2018). "Integration of Elbow Method and K-Means Algorithm for Clustering Analysis". Journal of Physics: Conference Series, Vol. 930, No. 1.

[36] Kodinariya, T. M., & Makwana, P. R. (2013). "Review on determining number of Cluster in K-Means Clustering". International Journal of Advance Research in Computer Science and Management Studies, 1(6), 90-95.

[37] Bholowalia, P., & Kumar, A. (2014). "EBK-Means: A Clustering Technique based on Elbow Method and K-Means in WSN". International Journal of Computer Applications, 105(9).

[38] Jain, A. K., & Dubes, R. C. (1988). *Algorithms for Clustering Data*. Prentice-Hall.

[39] Murtagh, F., & Contreras, P. (2012). Algorithms for hierarchical clustering: an overview. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 2*(1), 86–97.

[40] Ward Jr, J. H. (1963). Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association, 58*(301), 236–244.

[41] Nielsen, F. (2016). "Hierarchical Clustering". *Introduction to HPC with MPI for Data Science*, Springer, pp. 195-211.

[42] Manning, C. D., Raghavan, P., & Schütze, H. (2008). *Introduction to Information Retrieval*. Cambridge University Press.

[43] Jolliffe, I. T. (2002). *Principal Component Analysis*. Springer Series in Statistics.

[44] Shlens, J. (2014). "A tutorial on principal component analysis". *arXiv preprint arXiv:1404.1100*.

[45] Abdi, H., & Williams, L. J. (2010). "Principal component analysis". *Wiley Interdisciplinary Reviews: Computational Statistics*, 2(4), 433-459.

[46] Hotelling, H. (1933). "Analysis of a complex of statistical variables into principal components". *Journal of Educational Psychology*, 24(6), 417.

[47] Pearson, K. (1901). "On lines and planes of closest fit to systems of points in space". *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 2(11), 559-572.

[48] Cattell, R. B. (1966). "The scree test for the number of factors". *Multivariate Behavioral Research*, 1(2), 245-276.

[49] Pyle, D. (1999). *Data Preparation for Data Mining*. Morgan Kaufmann.

[50] Wirth, R., & Hipp, J. (2000). "CRISP-DM: Towards a standard process model for data mining". *Proceedings of the 4th International Conference on the Practical Applications of Knowledge Discovery and Data Mining*, pp. 29-39.

[51] Schafer, J. L. (1997). *Analysis of Incomplete Multivariate Data*. CRC Press.

[52] Allison, P. D. (2001). *Missing Data*. Sage Publications.

[53] Zheng, A., & Casari, A. (2018). *Feature Engineering for Machine Learning*. O'Reilly Media.

[54] Hughes, A. M. (2005). *Strategic Database Marketing*. McGraw-Hill Education.

[55] Mentzer, J. T., et al. (2001). *Logistics service quality as a segment-customized process*. Journal of Marketing

[56] Esper, T. L., et al. (2003). "The last mile: an examination of effects of online retail delivery strategies and customer satisfaction". *Journal of Business Logistics*.

[57] Aggarwal, C. C. (2017). *Outlier Analysis*. Springer Publishing Company.

[58] Hoaglin, D. C., & Iglewicz, B. (1987). "Fine-tuning some resistant rules for outlier labeling". *Journal of the American Statistical Association*.

[59] Iglewicz, B., & Hoaglin, D. C. (1993). *How to Detect and Handle Outliers*. ASQC Quality Press.

[60] Mohamad, I. B., & Usman, D. (2013). "Standardization and its effects on K-means clustering algorithm". *Research Journal of Applied Sciences, Engineering and Technology*, 6(17), 3299-3303.

[61] Pedregosa, F., et al. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research, 12*, 2825–2830.

[62] Shanker, M., Hu, M. Y., & Hung, M. S. (1996). "Effect of data standardization on neural network training". *Omega*, 24(4), 385-397.

[63] Milligan, G. W., & Cooper, M. C. (1988). "A study of standardization of variables in cluster analysis". *Journal of Classification*, 5(2), 181-204.

[64] Tan, P. N., Steinbach, M., & Kumar, V. (2016). *Introduction to Data Mining*. Pearson Education Limited.

[65] Jolliffe, I. T., & Cadima, J. (2016). "Principal component analysis: a review and recent developments". *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 374(2067).

[66] Payne, A., & Frow, P. (2013). *Strategic Customer Management: Integrating Relationship Marketing and CRM*. Cambridge University Press.

[67] Satish, G. N. V. R., et al. (2020). *An Optimal Cluster Selection using Elbow Method for Marketing Segmentation*. International Journal of Management and Analytics.

[68] Govindasamy, K., & Velmurugan, T. (2017). *Qualitative Performance Analysis of K-Means Clustering Algorithm in Supply Chain Management*. International Journal of Pure and Applied Mathematics.

[69] Wu, J. (2012). *Advances in K-means Clustering: A Data Mining Thinking*. Springer Science & Business Media.

[70] Li, Y., et al. (2021). *An extended RFM model for customer segmentation in the logistics industry*. Journal of Retailing and Consumer Services, 58.

[71] Moncrief, W. C. (2017). *Are sales as we know it dead?*. Journal of Personal Selling & Sales Management, 37(4).

[72] Hosseini, S. M., et al. (2010). *Mining customer value based on RFM model by K-means algorithm*. Journal of Applied Sciences, 10(21).

[73] Rousseeuw, P. J. (1987). *Silhouettes: A graphical aid to the interpretation and validation of cluster analysis*. Journal of Computational and Applied Mathematics.

[74] Kaufman, L., & Rousseeuw, P. J. (2009). *Finding Groups in Data: An Introduction to Cluster Analysis*. John Wiley & Sons.

[75] Hennig, C. (2015). *Clustering strategy and cluster validation*. Handbook of Cluster Analysis.

[76] Davies, D. L., & Bouldin, D. W. (1979). *A cluster separation measure*. IEEE Transactions on Pattern Analysis and Machine Intelligence.

[77] Caliński, T., & Harabasz, J. (1974). *A dendrite method for cluster analysis*. Communications in Statistics-theory and Methods.

[78] Xu, D., & Tian, Y. (2015). *A Comprehensive Survey of Clustering Algorithms*. Annals of Data Science.

[79] Handl, J., & Knowles, J. (2005). *Computational cluster validation in post-genomic data analysis*. Bioinformatics.

[80] Tipping, M. E., & Bishop, C. M. (1999). *Probabilistic principal component analysis*. Journal of the Royal Statistical Society: Series B (Statistical Methodology).

[81] Ringnér, M. (2008). *What is principal component analysis?*. Nature Biotechnology, 26(3), 303-304.

[82] Esper, T. L., et al. (2008). *Logistics service quality: A customer service perspective*. Journal of Business Logistics.

[83] Griffis, S. E., et al. (2012). *The impact of order fulfillment on customer retention and logistics service quality*. Journal of Business Logistics, 33(2).

[84] Kumar, V., & Reinartz, W. (2018). *Customer Relationship Management: Concept, Strategy, and Tools*. Springer.

[85] Ballou, R. H. (2007). *Business Logistics/Supply Chain Management*. Pearson Education India

EKLER

main.py

```python
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import datetime as dt
from sklearn.preprocessing import StandardScaler
from sklearn.cluster import KMeans, AgglomerativeClustering
from sklearn.metrics import silhouette_score, davies_bouldin_score, calinski_harabasz_score
from sklearn.decomposition import PCA

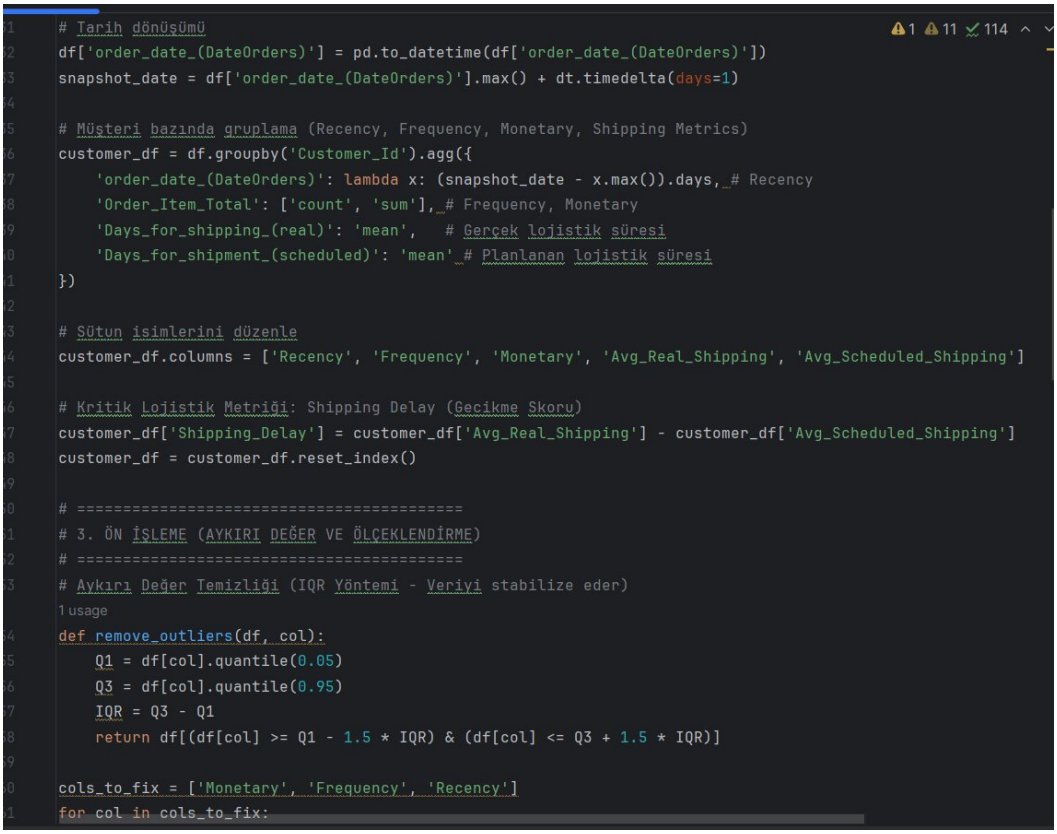
# ==========================================
# 1. VERİ YÜKLEME VE TEMİZLİK
# ==========================================
try:
    # Veri setini yükleme
    df = pd.read_csv( filepath_or_buffer: r'D:\Users\sedan\Downloads\archive (1)\DataCoSupplyChainDataset.csv', enc
    print("✅ Veri seti başarıyla yüklendi.")
except FileNotFoundError:
    print("❌ Hata: Dosya bulunamadı! Lütfen CSV dosyasının adını kontrol edin.")
    exit()

# Sütun isimlerindeki boşlukları temizle
df.columns = [col.replace(' ', '_') for col in df.columns]

# Eksik verileri temizle (Kritik alanlar boşsa satırı sil)
df = df.dropna(subset=['Customer_Id', 'Order_Item_Total', 'Days_for_shipping_(real)'])

# ==========================================
# 2. ÖZELLİK MÜHENDİSLİĞİ (RFM + LOJİSTİK)
# ==========================================
# Tarih dönüşümü
df['order_date_(DateOrders)'] = pd.to_datetime(df['order_date_(DateOrders)'])
```

Ek 1. Python Kaynak Kodları



```python
# Tarih dönüşümü
df['order_date_(DateOrders)'] = pd.to_datetime(df['order_date_(DateOrders)'])
snapshot_date = df['order_date_(DateOrders)'].max() + dt.timedelta(days=1)

# Müşteri bazında gruplama (Recency, Frequency, Monetary, Shipping Metrics)
customer_df = df.groupby('Customer_Id').agg({
    'order_date_(DateOrders)': lambda x: (snapshot_date - x.max()).days, # Recency
    'Order_Item_Total': ['count', 'sum'], # Frequency, Monetary
    'Days_for_shipping_(real)': 'mean',   # Gerçek lojistik süresi
    'Days_for_shipment_(scheduled)': 'mean' # Planlanan lojistik süresi
})

# Sütun isimlerini düzenle
customer_df.columns = ['Recency', 'Frequency', 'Monetary', 'Avg_Real_Shipping', 'Avg_Scheduled_Shipping']

# Kritik Lojistik Metriği: Shipping Delay (Gecikme Skoru)
customer_df['Shipping_Delay'] = customer_df['Avg_Real_Shipping'] - customer_df['Avg_Scheduled_Shipping']
customer_df = customer_df.reset_index()

# ========================================
# 3. ÖN İŞLEME (AYKIRI DEĞER VE ÖLÇEKLENDİRME)
# ========================================
# Aykırı Değer Temizliği (IQR Yöntemi - Veriyi stabilize eder)
1 usage
def remove_outliers(df, col):
    Q1 = df[col].quantile(0.05)
    Q3 = df[col].quantile(0.95)
    IQR = Q3 - Q1
    return df[(df[col] >= Q1 - 1.5 * IQR) & (df[col] <= Q3 + 1.5 * IQR)]

cols_to_fix = ['Monetary', 'Frequency', 'Recency']
for col in cols_to_fix:
```

Ek 2. Python Kaynak Kodları

```python
cols_to_fix = ['Monetary', 'Frequency', 'Recency']
for col in cols_to_fix:
    customer_df = remove_outliers(customer_df, col)

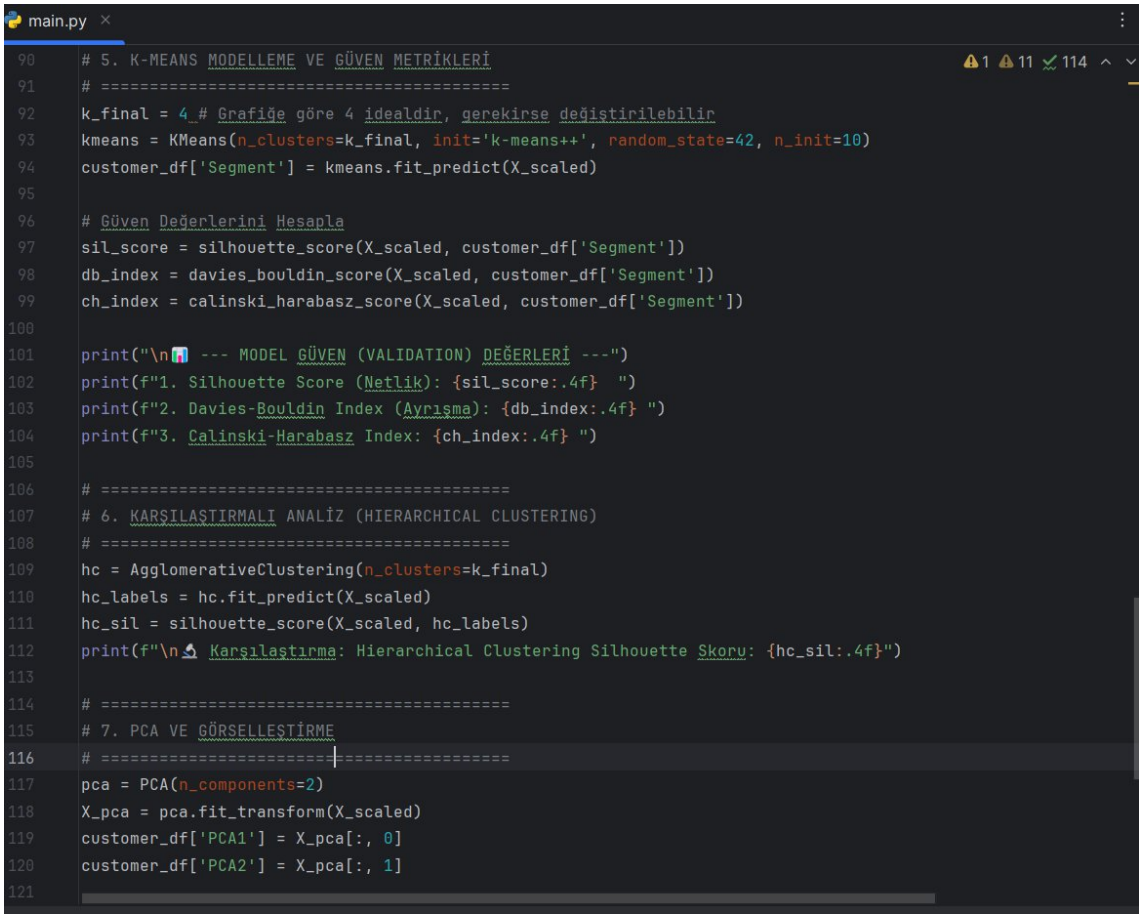
# Ölçeklendirme (K-Means için zorunludur)
features = ['Recency', 'Frequency', 'Monetary', 'Shipping_Delay']
X = customer_df[features]
scaler = StandardScaler()
X_scaled = scaler.fit_transform(X)

print(f"✅ Ön işleme tamamlandı. Kalan müşteri sayısı: {len(customer_df)}")

# ========================================
# 4. OPTİMUM KÜME SAYISI (ELBOW METHOD)
# ========================================
wcss = []
for i in range(1, 11):
    kmeans = KMeans(n_clusters=i, init='k-means++', random_state=42, n_init=10)
    kmeans.fit(X_scaled)
    wcss.append(kmeans.inertia_)

plt.figure(figsize=(8, 4))
plt.plot( *args: range(1, 11), wcss, marker='o', color='navy')
plt.title('Dirsek Metodu (Elbow Method)')
plt.xlabel('Küme Sayısı')
plt.ylabel('WCSS')
plt.grid(True)
plt.show()
```

Ek 3. Python Kaynak Kodları



```python
# 5. K-MEANS MODELLEME VE GÜVEN METRİKLERİ
# ==========================================
k_final = 4 # Grafiğe göre 4 idealdir, gerekirse değiştirilebilir
kmeans = KMeans(n_clusters=k_final, init='k-means++', random_state=42, n_init=10)
customer_df['Segment'] = kmeans.fit_predict(X_scaled)

# Güven Değerlerini Hesapla
sil_score = silhouette_score(X_scaled, customer_df['Segment'])
db_index = davies_bouldin_score(X_scaled, customer_df['Segment'])
ch_index = calinski_harabasz_score(X_scaled, customer_df['Segment'])

print("\n📊 --- MODEL GÜVEN (VALIDATION) DEĞERLERİ ---")
print(f"1. Silhouette Score (Netlik): {sil_score:.4f}  ")
print(f"2. Davies-Bouldin Index (Ayrışma): {db_index:.4f} ")
print(f"3. Calinski-Harabasz Index: {ch_index:.4f} ")

# ==========================================
# 6. KARŞILAŞTIRMALI ANALİZ (HIERARCHICAL CLUSTERING)
# ==========================================
hc = AgglomerativeClustering(n_clusters=k_final)
hc_labels = hc.fit_predict(X_scaled)
hc_sil = silhouette_score(X_scaled, hc_labels)
print(f"\n🔬 Karşılaştırma: Hierarchical Clustering Silhouette Skoru: {hc_sil:.4f}")

# ==========================================
# 7. PCA VE GÖRSELLEŞTİRME
# ==========================================
pca = PCA(n_components=2)
X_pca = pca.fit_transform(X_scaled)
customer_df['PCA1'] = X_pca[:, 0]
customer_df['PCA2'] = X_pca[:, 1]
```

Ek 4. Python Kaynak Kodları

```python
# ========================================
print("\n --- SEGMENT BAZLI İSTATİSTİKLER ---")


# 1. Ortalama değerler
summary_mean = customer_df.groupby('Segment')[features].mean().round(2)

# 2. Segment bazlı müşteri sayısı (order korunur)
segment_counts = customer_df.groupby('Segment')['Customer_Id'].count()

# 3. Yüzde hesaplama
segment_percent = (segment_counts / segment_counts.sum() * 100).round(2)

# 4. Tek tabloya birleştirme (index uyumlu şekilde)
summary_final = summary_mean.copy()
summary_final['Müşteri Sayısı'] = segment_counts
summary_final['Müşteri Yüzdesi (%)'] = segment_percent

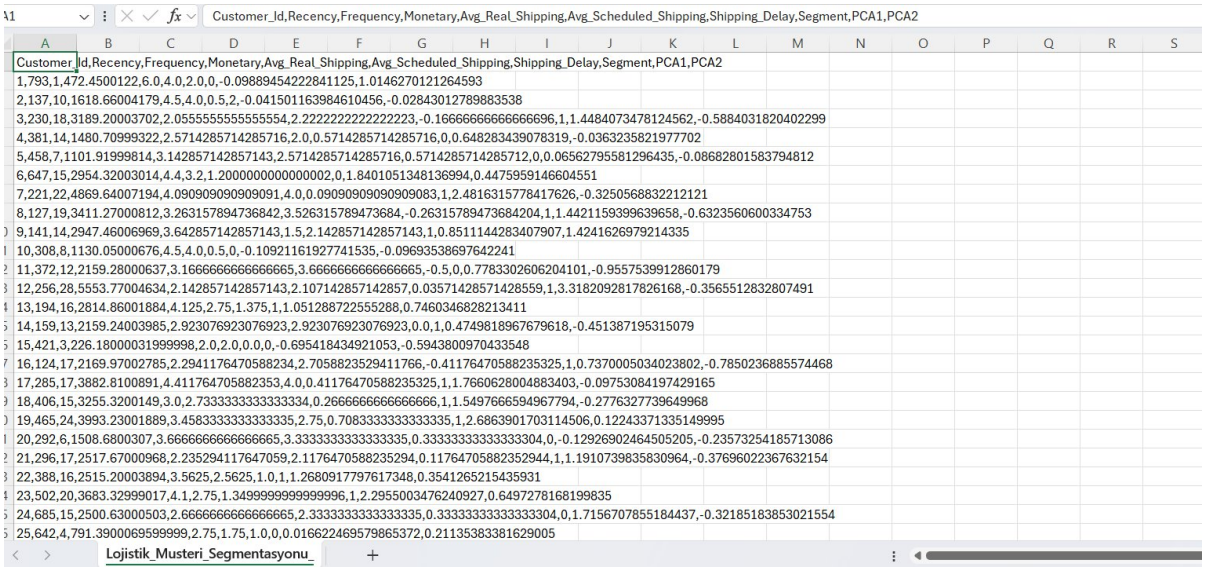
# 5. Segment sırasını garanti et
summary_final = summary_final.sort_index()

# 6. Sonucu yazdır
print(summary_final)

# 7. Tez için kaydet (UTF-8 BOM Excel uyumlu)
summary_final.to_csv(
    'Segment_Bazli_Ozet_Tablo.csv',
    encoding='utf-8-sig'
)

print("\n Segment bazlı özet tablo 'Segment_Bazli_Ozet_Tablo.csv' olarak kaydedildi ")
```

Ek 5. Python Kaynak Kodları



Customer_Id	Recency	Frequency	Monetary	Avg_Real_Shipping	Avg_Scheduled_Shipping	Shipping_Delay	Segment	PCA1	PCA2
1	793	1	472.4500122	6.0	4.0	2.0	0	-0.09889454222841125	1.014627012126459
2	137	10	1618.66004179	4.5	4.0	0.5	2	-0.041501163984610456	-0.02843012789883538
3	230	18	3189.20003702	2.0555555555555554	2.2222222222222223	-0.16666666666666696	1	1.4484073478124562	-0.5884031820402299
4	381	14	1480.70999322	2.5714285714285716	2.0	0.5714285714285716	0	0.648283439078319	-0.0363235821977702
5	458	7	1101.91999814	3.142857142857143	2.5714285714285716	0.5714285714285712	0	0.06562795581296435	-0.08682801583794812
6	647	15	2954.32003014	4.4	3.2	1.2000000000000002	0	1.8401051348136994	0.4475959146604551
7	221	22	4869.64007194	4.090909090909091	4.0	0.09090909090909083	1	2.4816315778417626	-0.3250568832212121
8	127	19	3411.27000812	3.263157894736842	3.526315789473684	-0.26315789473684204	1	1.4421159399639658	-0.6323560600334753
9	141	14	2947.46006969	3.642857142857143	1.5	2.142857142857143	1	0.8511144283407907	1.4241626979214335
10	308	8	1130.05000676	4.5	4.0	0.5	0	-0.10921161927741535	-0.09693538697642241
11	372	12	2159.28000637	3.1666666666666665	3.6666666666666665	-0.5	0	0.7783302606204101	-0.9557539912860179
12	256	28	5553.77004634	2.142857142857143	2.107142857142857	0.03571428571428559	1	3.3182092817826168	-0.3565512832807491
13	194	16	2814.86001884	4.125	2.75	1.375	1	1.051288722555288	0.7460346828213411
14	159	13	2159.24003985	2.923076923076923	2.923076923076923	0.0	1	0.4749818967679618	-0.451387195315079
15	421	3	226.18000031999998	2.0	2.0	0.0	0	-0.695418434921053	-0.5943800970433548
16	124	17	2169.97002785	2.2941176470588234	2.7058823529411766	-0.41176470588235325	1	0.7370005034023802	-0.7850236885574468
17	285	17	3882.8100891	4.411764705882353	4.0	0.4117647058823532	5	1.7660628004883403	-0.09753084197429165
18	406	15	3255.3200149	3.0	2.7333333333333334	0.2666666666666666	1	1.5497666594967794	-0.2776327739649968
19	465	24	3993.23001889	3.4583333333333335	2.75	0.7083333333333335	1	2.6863901703114506	0.12243371335149995
20	292	6	1508.6800307	3.6666666666666665	3.3333333333333335	0.33333333333333304	0	-0.1292690246450520	-0.23573254185713086
21	296	17	2517.67000968	2.235294117647059	2.1176470588235294	0.11764705882352944	1	1.1910739835830964	-0.37696022367632154
22	388	16	2515.20003894	3.5625	2.5625	1.0	1	1.2680917797617348	0.3541265215435931
23	502	20	3683.32999017	4.1	2.75	1.3499999999999996	1	2.2955003476240927	0.6497278168199835
24	685	15	2500.63000503	2.6666666666666665	2.3333333333333335	0.33333333333333304	0	1.7156707855184437	-0.32185183853021554
25	642	4	791.3900069599999	2.75	1.75	1.0	0	0.016622469579865372	0.21135383381629005

Lojistik_Musteri_Segmentasyonu_

Ek 6. Projenin Segmente Edilmiş Final Görseli

ÖZGEÇMİŞ

Adı Soyadı: Sedanur DURĞUN

Doğum Tarihi: 15.01.2002

Doğum Yeri: İstanbul

Lise: Sancaktepe Rabia Kız Anadolu İmamhatip Lisesi Fen ve Sosyal Bilimler Ağırlıklı Proje Okulu (2016-2020)

Staj Yaptığı Yerler: Uyumsoft Bilgi ve Teknolojileri A. Ş. (20 iş günü)