

T.C.
İSTANBUL AYDIN ÜNİVERSİTESİ
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ



PERAKENDE SEKTÖRÜNDE KÜMELEME ALGORİTMALARI
İLE MÜŞTERİ SEGMENTASYONU VE RFM MODELİ

YÜKSEK LİSANS TEZİ

Samet KANCA

Bilgisayar Mühendisliği Ana Bilim Dalı
Bilgisayar Mühendisliği Programı

EYLÜL, 2023

T.C.
İSTANBUL AYDIN ÜNİVERSİTESİ
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ



PERAKENDE SEKTÖRÜNDE KÜMELEME ALGORİTMALARI
İLE MÜŞTERİ SEGMENTASYONU VE RFM MODELİ

YÜKSEK LİSANS TEZİ

Samet KANCA
(Y2213.010019)

Bilgisayar Mühendisliği Ana Bilim Dalı
Bilgisayar Mühendisliği Programı

Tez Danışmanı: Doç. Dr. Yakup ÇELİKBİLEK

EYLÜL 2023

ONAY FORMU

ONUR SÖZÜ

Yüksek Lisans tezi olarak sunduğum “Perakende Sektöründe Kümeleme Algoritmaları İle Müşteri Segmentasyonu ve RFM Modeli” adlı çalışmanın, tezin proje safhasından sonuçlanmasına kadar ki bütün süreçlerde bilimsel ahlak ve etik geleneklere aykırı düşecek bir davranışımın olmadığını, tezdeki bütün bilgileri akademik ve etik kurallar içinde elde ettiğimi, bu tez çalışmasıyla elde edilmeyen bütün bilgi ve yorumlara kaynak gösterdiğimi ve yararlandığım eserlerin bibliyografyada gösterilenlerden oluştuğunu, bunlara atıf yaparak yararlanmış olduğumu belirtir ve onurumla beyan ederim.

Samet KANCA

ÖNSÖZ

Teknolojinin hızla ilerlediği ve dijital dönüşümün her sektörü etkisi altına aldığı bu dönemde, perakende sektörünün de bu dönüşümden nasıl etkilendiğini ve veri biliminin bu sektörde nasıl bir rol oynadığını araştırmak benim için büyük bir ilgi konusu olmuştur. Bu tez çalışması, bu ilgi alanımın bir ürünüdür ve perakende sektöründe müşteri segmentasyonunun önemini vurgulamayı amaçlamaktadır. Bu tez çalışması, perakende sektöründe müşteri segmentasyonunun önemini ve bu segmentasyonun işletmelere nasıl değer katabileceğini ortaya koymaktadır. Umarım bu çalışma, perakende sektöründe faaliyet gösteren işletmelere ve bu alanda araştırma yapmak isteyen diğer araştırmacılara katkıda bulunur.

Bu süreçte, bana rehberlik eden, bilgisi ve deneyimiyle her zaman yanımda olan hocam Doç. Dr. Tuncay Özcan ve tez danışmanı hocam Doç. Dr. Yakup Çelikkalek'e sonsuz teşekkürlerimi sunarım. Onların değerli katkıları olmadan bu çalışmayı tamamlamam mümkün olmazdı. Ayrıca, beni yetiştiren, hayatımın her aşamasında bana destek olan ve bugünlere gelmemde büyük emeği olan aileme de özverileri, sabırları ve inançları için en içten teşekkürlerimi sunarım.

Eylül, 2023

Samet KANCA

PERAKENDE SEKTÖRÜNDE KÜMELEME ALGORİTMALARI İLE MÜŞTERİ SEGMENTASYONU VE RFM MODELİ

ÖZET

Günümüzde teknolojinin gelişimi ile birçok sektörde olduğu gibi perakende sektöründe de gelişim ve dijital dönüşüm yaşanmaktadır. Bu çalışma, teknolojinin hızla geliştiği ve veri biliminin her sektörde önem kazandığı bir dönemde, perakende sektöründe müşteri segmentasyonunun önemini vurgulamaktadır. Özellikle tekstil perakendeciliği alanında, müşteri satın alma davranışlarının doğru bir şekilde analiz edilmesi ve segmente edilmesi, işletmelerin müşteri ilişkilerini yönetme ve pazarlama stratejilerini belirleme süreçlerinde kritik bir rol oynamaktadır. Bu çalışma, bir tekstil perakendecisinin maskelenmiş verilerini kullanarak, RFM modeli ile müşteri segmentasyonu yapmayı hedeflemektedir. Veriler, ön işleme tabi tutulmuş ve RFM değerleri hesaplanmıştır. Ardından, K-ortalama ve bulanık C-ortalama algoritmaları kullanılarak müşteri kümeleri oluşturulmuştur. Oluşturulan bu kümelerin sonuçları değerlendirilmiş ve müşteri gruplarına yönelik yorumlar yapmak için kullanılmıştır. Sonuçlar, müşteri segmentasyonunun, perakende sektöründe müşteri davranışlarını anlamak, müşteri ilişkilerini yönetmek ve etkili pazarlama stratejileri geliştirmek için ne kadar önemli olduğunu göstermektedir. Bu çalışma, perakende sektöründeki işletmelerin, müşteri verilerini etkin bir şekilde kullanarak, müşteri memnuniyetini artırmak ve işletme performansını iyileştirmek için nasıl stratejiler geliştirebileceğine dair değerli bilgiler sunmaktadır.

Anahtar Kelimeler: Perakende, RFM Analizi, Müşteri Segmentasyonu

CUSTOMER SEGMENTATION AND RFM ANALYSIS USING CLUSTERING ALGORITHMS IN RETAIL INDUSTRY

ABSTRACT

With the development of technology, as in many sectors, the retail sector is also experiencing development and digital transformation. This study emphasizes the importance of customer segmentation in the retail sector in a period where technology is rapidly developing, and data science is gaining importance in every sector. Especially in the field of textile retailing, the correct analysis and segmentation of customer purchasing behaviors play a critical role in managing customer relationships and determining marketing strategies for businesses. This study aims to perform customer segmentation using the masked data of a textile retailer with the RFM model. The data has been preprocessed and RFM values have been calculated. Then, customer clusters were created using K-means and Fuzzy C-means algorithms. These clusters were evaluated to make comments on customer groups. The results show how important customer segmentation is to understand customer behaviors in the retail sector, manage customer relationships, and develop effective marketing strategies. This study provides valuable information on how businesses in the retail sector can develop strategies to increase customer satisfaction and improve business performance by effectively using customer data.

Keywords: Retail, RFM Analysis, Customer Segmentation

İÇİNDEKİLER

	<u>Sayfa</u>
ONUR SÖZÜ	i
ÖNSÖZ.....	ii
ÖZET.....	iii
ABSTRACT	iv
İÇİNDEKİLER	v
KISALTMALAR LİSTESİ.....	vii
ŞEKİLLER LİSTESİ.....	viii
I. GİRİŞ	1
A. Problem Tanımı	2
B. Proje Hedefi	2
C. Önerilen Çözüm.....	2
D. Proje Kapsamı.....	3
II. LİTERATÜR İNCELEME	5
III. METODOLOJİ.....	20
A. Makine Öğrenmesi.....	20
1. Denetimli Öğrenme.....	21
2. Denetimsiz Öğrenme	22
B. Kümeleme.....	22
1. K-Ortalama	22
2. Bulanık C-Ortalama	24
C. Değerlendirme Yöntemleri	26

1. Silhouette indeksi.....	26
D. RFM Analizi	26
1. Recency.....	28
2. Frequency.....	28
3. Monetary	29
IV. UYGULAMA.....	30
A. Veri Seti ve Veri Ön İşleme.....	31
B. Kümeleme Algoritmalarının Uygulanması.....	34
1. K-ortalama ile RFM Analizi	37
2. Bulanık C-Ortalama ile RFM Analizi.....	43
V. SONUÇ.....	48
VI. KAYNAKÇA	50
EKLER.....	56
ÖZGEÇMİŞ.....	67

KISALTMALAR LİSTESİ

AI	: Artificial intelligence – Yapay Zeka
CART	: Classification and Regression Tree
CHAID	: Chi Squared Automatic Interaction Detection - Otomatik Ki-Kare Etkileşim Belirleme Analizi
CRM	: Customer Relationship Management – Müşteri İlişkileri Yönetimi
DBSCAN	: Density-Based Spatial Clustering of Applications with Noise
FCM	: Bulanık C-Ortalamalar Kümeleme (Fuzzy C-Means Clustering)
K-Means	: K-Ortalama Kümeleme
LTV	: Customer Lifetime Value - Müşteri Yaşam Boyu Değer
PESTEL	: Politik, Ekonomik, Sosyal, Teknolojik, Yasal ve Çevre
RFM	: Recency (Sıklık), frequency (Frekans) ve monetary (Harcama Tutarı)
TÜİK	: Türkiye İstatistik Kurumu

ŞEKİLLER LİSTESİ

	<u>Sayfa</u>
Şekil 1. RFM Literatür Taraması Özet Tablo.....	19
Şekil 2. Firmadan alınan ham müşteri satış verileri.	32
Şekil 3. Müşteri ham verisine ait kolon isimleri veri açıklamaları.....	32
Şekil 4. Ön işlemeden geçirildikten sonra müşteri veri deseni.	33
Şekil 5. Sipariş sayısının müşteriye oranı.....	35
Şekil 6. K-ortalama ve bulanık c-ortalama algoritmalarının küme başarımlarının karşılaştırılması	35
Şekil 7. K-ortalama algoritmasının farklı küme sayılarının toplam küme içi uzaklıkları.....	36
Şekil 8. Bulanık C-ortalama algoritmasının farklı küme sayılarının toplam küme içi uzaklıkları.....	37
Şekil 9. K-ortalama algoritması sekizli küme sonuçları.....	38
Şekil 10. Sıklık ve harcama tutarı değerlerinin küme dağılımı	42
Şekil 11. Frekans ve harcama tutarı değerlerinin küme dağılımı.....	42
Şekil 12. Bulanık C-ortalama küme sonuçları.....	43
Şekil 13. Bulanık C-ortalama kümelerinin üç boyutlu şekillendirmesi	45
Şekil 14. Sıklık ve harcama miktarı değerlerinin küme dağılımı.....	46
Şekil 15. Frekans ve harcama tutarı değerlerinin küme dağılımı.....	47

I. GİRİŞ

Müşteri ilişkileri yönetimi, bir şirketin müşterileriyle derinlemesine ve karşılıklı faydalı ilişkiler kurma kapasitesini yansıtan kritik bir işlemdir. Bu, bir şirketin tüketicilerle nasıl etkileşimde bulunduğunu, hangi stratejileri benimsediğini ve hangi prosedürleri takip ettiğini belirtir. Müşterinin bir ürünü keşfetme, satın alma ve sonrasında yaşadığı deneyimin her aşaması, müşteri ilişkileri yönetiminin kapsamı içindedir.

RFM analizi, müşteri segmentasyonunu daha verimli ve hedef odaklı hale getiren bir araçtır. Bu analiz, şirketlere, ürünleriyle en aktif etkileşimde bulunan ve satın alma olasılığı en yüksek olan tüketicilere nasıl ulaşacakları konusunda rehberlik eder. RFM'nin müşteri segmentasyon sürecinde merkezi bir role sahip olması, firmaların, müşterilerin geçmişteki ve mevcut satın alma davranışlarına daha duyarlı stratejilerle yanıt vermesini sağlar. Bu, pazarlama kampanyalarının etkinliğini artırarak müşteri memnuniyetini yükseltir. RFM segmentasyonunun gücü, şirketlere, özellikle dikkate alınması gereken ve potansiyel olarak en yüksek getiriyi sağlayabilecek müşteri gruplarını tanımlama yeteneği sunar. Bu, özellikle kaynakların en etkili şekilde nasıl tahsis edileceğine dair stratejik kararlar alırken kritik öneme sahiptir. Ayrıca, müşterilerden gelecek olan toplam nakit akışının şu anki değerini analiz ederek, şirketlerin hangi müşteri segmentlerinin en değerli olduğunu belirlemelerine yardımcı olur. Bu, her bir müşteri grubu için özelleştirilmiş iş ve pazarlama stratejileri oluşturulmasını ve uygulanmasını teşvik eder. Bununla birlikte RFM analizi ile uzun bir dönemdir alışveriş yapmayan müşteriler tespit edilip bu müşteriler özelinde de pazarlama ve tekrardan firmadan ürün almaya yönelik faaliyetler yürütülebilir. Uzun vadede, bu yaklaşımlar bütünü şirketlerin odaklarını sadece kısa vadeli karlardan, müşterilerle sürdürülebilir ve uzun vadeli ilişkilere doğru kaydırmalarına yardımcı olabilir. Bu hem müşteri sadakatini artırır hem de şirketin sürdürülebilir büyümesini destekler.

A. Problem Tanımı

Son yıllarda, teknolojideki hızlı gelişmeler, ürün çeşitliliğindeki büyük artış ve ürün kalitesindeki sürekli iyileştirmeler, perakende sektöründeki liderlik standartlarını ve beklentilerini önemli ölçüde değiştirdi. Bu evrimsel süreç, müşteri beklentilerinde de derinlemesine bir dönüşüme yol açtı. Daha önceleri, müşteriler genellikle ürünün temel özelliklerine ve fiyatına odaklanıyordu. Ancak günümüzde, modern tüketiciler sadece ürünün kendisine değil, aynı zamanda satın alma deneyimine, markanın değerlerine, ürünün veya hizmetin onların benzersiz ihtiyaçlarına ne kadar uygun olduğuna da bakıyorlar. Bu nedenle, şirketlerin tüketicilere aynı ürünleri sunarak ve onları tek bir kategoriye koyarak genelleme yapmaları artık kabul edilemez. Her bireyin kendi benzersiz ihtiyaçların ve beklentilerinin olduğunu kabul etmeleri gerekmektedir. Eğer bir şirket bu benzersiz ihtiyaçları ve beklentileri anlamaz veya karşılamazsa, müşteri, ihtiyaçlarını daha iyi anlayan ve karşılayan başka bir şirkete yönelebilir.

Bu bağlamda, şirketlerin müşteri ilişkileri yönetiminde proaktif olmaları, müşteri geri bildirimlerini dikkate almaları ve ürün veya hizmetlerini sürekli olarak müşteri ihtiyaçlarına göre uyarlamaları esastır. Bu, sadece müşteri memnuniyetini artırmakla kalmaz, aynı zamanda marka sadakatini de güçlendirir. Sonuç olarak, perakendecilerin eski, tek tip yaklaşımlarını terk edip, daha bireysel, kişiselleştirilmiş stratejilere yönelmeleri gerekmektedir. Bu hem şirketlerin hem de müşterilerin yararına olacak bir dönüşümdür.

B. Proje Hedefi

Yapılan bu çalışma ile çok ürünü ve çok lokasyonu olan perakende firmalarının geçmiş veriyi kullanarak geleceğe yönelik çıkarımlarını daha doğru yapmalarını sağlayacak ve geçmiş verideki özel durumları algılayabilecek bir uygulama. Geçmiş veri setini bozan bilinmeyen özel durumlar (perakende takvimi, bayram, özel gün vb. dışında gelişen) minimize edilecektir.

C. Önerilen Çözüm

Hedef kitleye odaklanma stratejisi, işletmelere, en yüksek satın alma potansiyeline sahip olan müşterilere pazarlama faaliyetlerini yoğunlaştırma fırsatı

sunar. Müşteri segmentasyonu, işletmelere hedefledikleri kitleyi daha detaylı ve net bir şekilde tanımalarına olanak tanır. Bu projenin amacı, perakende sektöründe, özellikle tekstil alanında faaliyet gösteren işletmeler için daha etkili ve hedef odaklı müşteri stratejileri oluşturmaktır.

Bu proje kapsamında, bir tekstil perakendecisinin müşteri verilerini detaylı şekilde inceleyerek, bu müşterileri belirli özelliklerine göre farklı segmentlere ayıracağız. Segmentasyon sürecinde RFM analizinin yanı sıra K-ortalama ve bulanık C-ortalama algoritmalarını kullanarak daha keskin ve anlamlı müşteri grupları oluşturmayı hedefliyoruz. Oluşturduğumuz bu segmentler üzerinde derinlemesine analizler gerçekleştirip, elde edilen bulguları detaylı bir şekilde yorumlayacağız. Bu analizlerin sonucunda, firmanın hangi müşteri grubuna hangi stratejilerle yaklaşması gerektiğine dair somut önerilerde bulunacağız. Bu, firmanın hem mevcut müşteri ilişkilerini güçlendirmesine yardımcı olacak hem de potansiyel müşterilere nasıl daha etkili bir şekilde ulaşabileceği konusunda rehberlik edecektir. Sonuç olarak, bu proje ile firmanın müşteri deneyimini zenginleştirilmesi ve müşteri memnuniyetini artırması yönünde alacağı kararlara katkılar sağlamayı hedefliyoruz.

D. Proje Kapsamı

Bu proje kapsamında, RFM analizi kullanılarak gerçekleştirilecek müşteri segmentasyonu için belirlenen adımlar şu şekildedir:

1. **Veri Ön İşleme:** Müşteri verileri detaylı bir şekilde incelenecek, eksik veya hatalı veri girişleri tespit edilip düzeltilerek veri seti temizlenecektir. Anomalilerin ayıklanması ve verinin bütünlüğünün artırılması, analizin doğruluğunu ve güvenilirliğini arttıracaktır.
2. **RFM Değerlerinin Oluşturulması:** Tekil müşteri bazında Sıklık (Frequency), Sonluk (Recency) ve Parasal Değer (Monetary) olmak üzere RFM değerleri hesaplanacak ve her müşteri için bu değerler belirlenecektir.
3. **Kümeleme Algoritmalarının Uygulanması:** Müşteri verileri, K-ortalama ve bulanık C-ortalama algoritmaları kullanılarak segmentlere ayrılacaktır.

Bu iki algoritma, segmentasyon için farklı yaklaşımlar sunarak daha geniş bir perspektif kazandıracaktır.

4. **Başarım Değerlendirmesi:** Oluşturulan kümelerin başarımı, Silhouette metriği ile değerlendirilecektir. Bu değerlendirme, hangi algoritmanın daha tutarlı ve anlamlı segmentler oluşturduğunu belirlememize yardımcı olacaktır. En yüksek Silhouette değerine sahip olan kümeleme yöntemi, en başarılı olarak kabul edilecektir.
5. **Küme Değerlendirme ve Yorumlama:** Son adımda, müşterilerin satın alma geçmişlerine dayanarak oluşturulan kümeler detaylı bir şekilde incelenecek ve yorumlanacaktır. Bu yorumlama, her bir segmentin özelliklerini, davranışlarını ve potansiyel ihtiyaçlarını anlamamıza yardımcı olacaktır.

Bu adımların tamamlanmasının ardından, işletmelere müşteri tabanlarını daha etkili bir şekilde yönetme ve pazarlama stratejilerini optimize etme konusunda rehberlik edecek değerli bilgiler sunulmuş olacaktır.

II. LİTERATÜR İNCELEME

Perakende sektörünün yapısı her zamankinden daha hızlı bir şekilde evrilmektedir. Hem yerel hem de çok uluslu işletmelerden artan bir rekabet, şirket birleşmeleri, yeni edinimler ve değişimler, alışveriş deneyimleri konusunda beklentileri olan daha donanımlı ve zorlu müşteriler gözlemlemeye başladık (Dabholkar ve diğ., 1996). Bunun doğal bir sonucu olarak, perakendeciler bugün rakiplerinden daha üstün bir hizmet sağlama konusunda kendilerini çeşitlendirecek yeni bir stratejik yönlendirmeyi benimsemelidir. Bu genellikle rekabette öne çıkmanın en etkili perakende stratejilerinden biri olarak ifade edilir (Reichheld ve Sasser, 1990).

Birçok şirket, müşteri talepleri ve ihtiyaçları hakkında bilgi edinmek için Müşteri İlişkileri Yönetimi gibi, şirketin rekabetçi pazarını güçlendirecek bilgileri maksimum düzeyde kullanamamaktadır. Ayrıca, kullanıcı verilerini madencilik ve hesaplamalı yöntemlerin kombinasyonu kullanılmış olmasına rağmen, bu yaklaşımların bazı sınırlılıkları vardır. Ayrıca, sadece birkaç çalışma kullanıcıyı anlama konusunda RFM modelini kullanmaktadır. ([Bachtiar](#), 2018)

Endüstriyel sektörde pazarlama, maliyetli bir faaliyet olduğundan doğru bir şekilde gerçekleştirilmelidir. Müşteri satış verilerinin analizi, bilinen bir yöntem olan Yakınlık, Sıklık ve Parasal Değer (RFM) ile veri analitiği tekniğinin birleştirilmesiyle yapılmıştır. Sonuçlar, veri analitiğinin, kar getirmeyen müşterileri dikkate alarak geleneksel RFM'nin doğruluğunu artırabileceğini ve gerçek zamanlı veri sağlayabileceğini göstermektedir. Bu araştırma, pazarlamacılar için basit ve yaygın olarak kullanılan bir araç olan geleneksel RFM analizinin etkinliğini artırma konusunda yeni bir bakış açısı sunmaktadır. (Purnomo ve diğ., 2021)

Pazarlama çalışmaları, yoğun rekabet ortamında ayakta kalmayı hedefleyen işletmeler açısından müşterilerin önemine sıklıkla dikkat çekmiştir. Müşteri İlişkileri Yönetimi, müşterilerle ilişkileri geliştirmeyi amaçlayan öne çıkan bir pazarlama yönetimi yaklaşımı olmuştur. Müşteri ilişkileri yönetimi yaklaşımının

pratik bir uygulaması, işletmeler ve müşteriler için değer çıkarmak üzere müşteri verilerinin analizidir. Bu bağlamda, müşteri segmentasyonu, benzer niteliklere sahip müşteri gruplarını belirlemeye ve müşteri grupları için daha iyi uyarlanmış pazarlama stratejilerini belirlemeye yardımcı olan bir görev olmuştur. Müşteri segmentasyonu için ortaya konulmuş çeşitli yaklaşımlar arasında RFM Modeli, etkin ve kolay uyarlanabilir olmasıyla öne çıkmaktadır. Müşterilerin satış verisine ilişkin 3 farklı boyut üzerinden sıralanmasına dayanan yöntem, sıralamada kullanılan puanlama biçimine göre çeşitli yaklaşımlara konu olmaktadır (Köse, U., ve Arslan, A. 2020).

Pazarlamacıların çoğu, başarılı kampanyalar için doğru müşterileri belirleme konusunda zorluk çeker. Şimdiye kadar, bir lansman kampanyası için uygun müşterileri seçmek için kullanılan popüler bir yöntem müşteri segmentasyonudur. Ne yazık ki, müşteri segmentasyonu ve pazarlama kampanyası arasındaki bağlantı eksiktir. Başka bir problem, veritabanı pazarlamacılarının genellikle müşteri segmentasyonunu ve müşteri hedeflemeyi gerçekleştirmek için farklı modeller kullanmasıdır. Chan tarafından yapılan bir çalışmada kampanya stratejileri için müşteri hedeflemeyi ve müşteri segmentasyonunu birleştiren yeni bir yaklaşım sunmaktadır. Chan araştırmasında RFM modeli kullanarak müşteri davranışını belirler ve ardından önerilen segmente edilmiş müşterileri değerlendirmek için bir Müşteri Yaşam Boyu Değer modeli kullanmıştır. Önerilen yöntemin verimliliğini göstermek için, bu çalışma, 4000'den fazla müşteriyi segmente etmek için bir Nissan otomobil perakendecisinin ampirik bir çalışmasını gerçekleştirmektedir. Deneysel sonuçlar, önerilen yöntemin değerli müşterileri rastgele seçime göre daha etkili bir şekilde hedefleyebileceğini göstermektedir. (Chan, 2008)

Cheng ve Chen tarafından yapılan bir çalışma (2009), RFM bileşenlerini ve K-ortalama kümeleme algoritmasını teorilere aktaran bir süreç önermektedir. Bu, sadece sınıflandırma sürecinin doğruluk seviyesini artırmak için değil, aynı zamanda şirketler için üstün bir müşteri ilişkileri yönetimi oluşturmak için bir sınıflandırma kural listesi elde etmek için de tamamlanmıştır. Önerilen model ayrıca veri madenciliği teknolojilerinin çoğu olumsuz etkisini başarıyla hafifletmektedir. Çalışmanın sonunda, bu çalışmanın şirketlerin hedef müşterilere odaklanmasına ve ardından hem şirket hem de müşteri için olumlu bir durum

yaratarak gelirleri maksimize etmesine yardımcı olacağına inanılmaktadır (Cheng ve Chen, 2009).

Kuo ve Cheng tarafından yapılan bir araştırmada (2009), iki yeni kriter olan güncellik ve parasal değer tanıtılmıştır. Bu değerler güncellik, sıklık ve parasal ağırlık, çapraz satış, ürün önerisi ve kişiselleştirilmiş pazarlama dahil olmak üzere çeşitli e-ticaret uygulamalarında kullanılabilir olduğunu göstermiştir. (Kuo ve Cheng, 2009).

RFM modeli, araştırmacılara ve karar vericilere yararlı bilgiler sağlayabilecek oldukça etkili bir modeldir. RFM modeli, çeşitli durumlarda son derece etkili olduğu kanıtlanmıştır. Wei, Lin ve Wu tarafından yapılan bir araştırmada (2010), RFM sadece kâr amacı güden firmalara değil, aynı zamanda değerli müşterileri belirleme ve etkili pazarlama stratejileri geliştirme konusunda kâr amacı gütmeyen kuruluşlara ve hükümet ajanslarına da yardımcı olabileceğini göstermiştir (Wei, ve diğ., 2010).

Teknolojinin ilerlemesi ve İnternet'in yaygınlaşmasıyla, adını sıkça duymaya başladığımız büyük veri kavramı ortaya çıkmıştır. Büyük veri, yapılandırılmamış veri yığını olarak kısaca tanımlanabilir. Farklı kaynaklardan toplanan veriyi anlamlı ve işlenebilir bir formata dönüştürmeyi amaçlar. Bu yöntemlerden biri RFM analizidir. Erpolat ve arkadaşları tarafından yapılan çalışmada, RFM analizinin önemine vurgu yapılmıştır. RFM analizinin tahmin modellerinde nasıl kullanılacağı detaylı olarak açıklanmıştır. RFM analizinin geri dönüşüm projesine uygulanabilirliği gösterilmiştir. Modelin işleyişi ve RFM'nin geri dönüşüm projelerine uygulanması, metal sektöründe faaliyet gösteren bir şirketin orijinal perakende verilerinde gösterilmiştir. Bu sayede, geri dönüşüme katılımı teşvik edilmiştir. Geri dönüşümün katkı oranına göre, müşterilere indirimler uygulayarak müşteri ve şirket arasında karlı bir ilişki kurulması hedeflenmiştir. (Erpolat ve diğ., 2020)

Günümüz iş ortamında müşteri sermayesine yönelik artan önem göz önüne alındığında, birçok firma müşteri sadakati ve karlılık kavramlarına odaklanmaktadır. Firma, daha hedeflenmiş ve kişiselleştirilmiş pazarlama stratejilerini oluşturmak için gerekli olan temel bilgileri sağlayabilecek müşteri değeri konseptinden yola çıkarak, başarılı bir Müşteri İlişkileri Yönetimi

oluşturmayı amaçlar. Bir sağlık ve güzellik şirketinde uygulanan bir çalışma, RFM pazarlama analiz yönteminin yanı sıra, 'Count Item' adı verilen ek bir parametre içeren genişletilmiş RFM analiz yönteminin müşteri segmentasyonunda kullanıldığını göstermiştir. Her iki yaklaşımın sonuçları karşılaştırıldığında, 'Count Item'ın RFM yöntemine eklenmesinin kümeleme sonucunda belirgin bir fark yaratmadığı, bu nedenle Müşteri Yaşam Boyu Değeri'nin her segment için ağırlıklı RFM yöntemi temelinde hesaplandığı görülmüştür. Bu hesaplamalar, şirketin pazarlama ve satış stratejilerini belirlemesine yardımcı olabilir (Khajvand ve Zolfaghar, 2010).

Kim tarafından yapılan çalışmada, RFM analizinin gelecekteki çekirdek teknolojileri tahmin etmek için etkin bir şekilde kullanılabileceğini göstermektedir. Amerika Birleşik Devletleri patent verileri kullanılarak elde edilen deneysel sonuçlar, yakınlık, sıklık ve parasal değer, gelecekteki çekirdek teknolojisi patentlerini belirlemek için etkili değişkenler olduğunu göstermektedir. Ayrıca, sınıflandırma ve regresyon ağacı ile birleştirilen iki örnekleme kullanılarak gelecekteki çekirdek teknolojiyi belirlemek için kurallar aranmış ve öğrenme algoritmaları doğruluk, geri çağırma ve F-ölçüsü açısından karşılaştırılmıştır. Hesaplamalı çalışmalar, aşırı örnekleme yönteminin, gelecekteki çekirdek teknolojiyi belirlemek için veriler gibi dengesiz verilerden kuralları bulmak için etkili olduğunu göstermektedir (Kim ve diğ., 2012).

RFM analizi üzerine yapılan başka bir çalışmada segmentasyon gerçekleştirilmiş ve sonrasında K-ortalama kümeleme, bulanık C-ortalama ve RM K-ortalama algoritmaları kullanılarak müşteriler kümelendi. Bu algoritmaların çalışma süreleri ve etkinlikleri analiz edilmiştir. Sonuç olarak, RM K-Means algoritmasının diğer tekniklere göre daha az zaman aldığı ve iterasyon sayısını azalttığı gözlemlenmiştir (Christy, 2018).

Kişiselleştirilmiş müşteri hizmeti sunan hassas pazarlama, yüksek verimli pazarlama yöntemleri aracılığıyla işletmelerin karlarını artırmalarına yardımcı olur. You ve arkadaşları tarafından yapılan çalışmada, veri madenciliği tekniklerini kullanarak hassas pazarlamaya yönelik yeni bir karar verme çerçevesi sunulmaktadır. İlk olarak, bu çalışma aylık tedarik miktarını doğru bir şekilde tahmin etmek için bir trend modeli sunar; ikinci olarak, RFM modelini, müşterileri farklı gruplara kümelenecek için öznitelik seçmek amacıyla kullanır;

üçüncü olarak, Otomatik Ki-Kare Etkileşim Belirleme Analizi karar ağaçları ve Pareto değerlerini, farklı müşteri gruplarını ayırt etmek için önemli öznelik değerlerini belirlemek için kullanır; ve son olarak, RFM ile ayrılmış her müşteri grubuna yönelik farklı tedarik stratejileri oluşturur. Önerilen hassas karar verme çerçevesinin amacı, yöneticilere farklı müşteri kategorilerinin potansiyel özelliklerini belirlemeye ve uygun hassas pazarlama stratejileri önermeye yardımcı olmaktadır. Bu çalışmada Çin'deki bir şirketten gerçek dünya verileri toplanmış ve önerilen çerçevenin nasıl uygulanacağını göstermek için bir durum çalışması olarak kullanılmıştır. Bu durum çalışması, yapılan çalışmanın hassas pazarlama stratejisi sağlama yeteneğine sahip olduğunu göstermektedir (You, ve diğ., 2015).

Müşterinin yaşam süresi, organizasyonun müşterisinden daha fazla kar elde edebileceği müşteri ve organizasyon arasındaki ilişkiye işaret eder. RFM modeli, yakınlık, sıklık ve parasal değer olmak üzere üç değişkene dayanmaktadır. Nikumanesh ve Albadvi tarafından yapılan bu vaka çalışmasında (2014), İran Tarım Bankası'nın veritabanlarını inceleyip, yapılan anketler kullanarak uzmanlarla görüşmeleri içermektedir. Toplanan bu veriler ile yapılan RFM modelinin parametreleri, yakınlık, sıklık ve parasal değer olmak üzere üç alanda beş değişkene sınırlıdır. Değişkenler, uzmanların yargılarına göre ağırlıklandırılır. Daha sonra, 12.359 müşteri veri kaydı üzerinde yapılan bir kümeleme yöntemi ile müşteriler 852 özel, 1755 ana ve 9373 normal sınıfa ayrılır. Ayrıca, müşteriler modelin üç parametresinin ortalamasına olan mesafeye göre sekiz gruba ayrılır. Sonuçlar, toplam müşterilerin %80'ini oluşturan yani normal müşterilerin en düşük karlılığa sahip olduğunu göstermektedir (Nikumanesh ve Albadvi, 2014).

Şirketler, veri kalitesindeki problemlerin segmentasyon tekniklerinin performansını nasıl etkilediği ve hangi tekniklerin bu problemlere diğerlerinden daha dayanıklı olduğunu bilmekten büyük ölçüde faydalanır. Coussement ve arkadaşları tarafından yapılan bu çalışma (2014), direkt pazarlama için üç önemli segmentasyon tekniği olan RFM analizi, lojistik regresyon ve karar ağaçları üzerindeki etkisini inceler. Analiz edilen iki gerçek hayat direkt pazarlama veri seti için, sonuçlar; (1) optimal veri doğruluğu altında, karar ağaçlarının RFM analizi ve lojistik regresyondan üstün olduğunu, (2) veri doğruluğu problemlerinin girişinin, üç segmentasyon tekniğinin performansını bozduğunu

ve (3) verilerin daha az doğru olduđu durumlarda, karar ağaçlarının lojistik regresyon ve RFM analizinden üstün olduğunu gösterir. Genel olarak, bu çalışma, veri doğruluk problemlerinin şüphesi altında bile, direkt pazarlama için müşteri segmentasyonu bağlamında karar ağaçlarının kullanılmasını önermektedir (Coussement ve diğ., 2014).

Bilgi teknolojisinin büyük bir gelişme göstermesi, veri miktarında bir patlamaya neden olmuştur, ancak bu verinin faydalı içgörüler elde etmek için işlenmesi gerekmektedir. Müşteri İlişkileri Yönetimi olarak da adlandırılan daha iyi ilişkiler kurmak için müşterinin ihtiyaçlarını, davranışlarını ve değerlerini incelemek için verinin kullanılması gerekmektedir. Şirket büyüdükçe, veri artmakta ve müşterilerle doğrudan etkileşimde bulunmak daha zor hale gelmektedir. Bu nedenle, daha az etkili pazarlama kampanyaları gibi sorunlar, hemen ele alınmazsa zararlara neden olabilir. Bu bağlamda, RFM değişkenleri kullanılarak ve dirsek yöntemi ve siluet skoru ile küme sayısını belirleyerek K-ortalama kümeleme ile müşteri segmentasyonu gerçekleştirilmiştir. Analiz sonuçlarına göre, en iyi müşteriler, kaybedilmeyebilecek müşteriler ve ortalama müşteriler olmak üzere üç tür küme bulunmuştur. (Barus, ve diğ., 2023)

Sarvari ve arkadaşları tarafından yapılan çalışmada (2016), müşteri segmentasyonuna en iyi yaklaşımın ne olduğunu belirlemek ve RFM ve demografik faktörlere dayalı kuralları öngörmek, olduğu görülmüştür. Bu çalışmada farklı senaryolar tasarlanmış, uygulanmış ve tek tip test koşulları altında titizlikle değerlendirildiği görülmüştür. Ana fikir, müşteri kümelerinin, demografik verilerle birlikte RFM analizi temelindeki segmentasyon süreçlerine dayandığında daha iyi geliştirildiğidir. Ağırlıklı RFM ve ağırlıksız RFM değerleri/puanları, demografik faktörlerle ve demografik faktörler olmadan uygulanmış ve farklı tür ve sayıda küme oluşturmak için kullanılmıştır. Sonuçlar, uygun bir segmentasyon yaklaşımının olmasının, güçlü ilişkilendirme kuralları için hayati öneme sahip olduğunu göstermiştir. Ayrıca, daha doğru müşteri segmentlerini yakalamak için, kümeleme, RFM ve demografik özelliklerin kombinasyonu önerilmiştir. Bu çalışma, demografik verilerin ağırlıklı RFM ile birleştirilmesinin önemini vurgulamaktadır (Sarvari, ve diğ., 2016).

RFM Analizi ile özelinde yapılan bir çalışmada, klasik RFM modeline PESTEL (Political, Economic, Social, Technological, Legal ve Environmental)

analizinin bir bileşeni olan "ekonomik" değişkeni eklenerek RFMS (Recency, Frequency, Monetary, Sensitivity) modeli önerilmiştir. Bu model, şirketlerin verimlilik süreçlerini iyileştirmeyi ve müşteri ilişkilerini daha doğru yönetecek bir sınıflandırma yöntemi geliştirmeyi amaçlamaktadır. Böylece, zaman ve maliyet tasarrufu sağlanarak müşteri ile şirket arasında karlı bir ilişki kurulması hedeflenmiştir. Önerilen modelin etkileri, BORUSANCAT Makine ve Güç Sistemleri'nin müşteri veritabanı kullanılarak analiz edilmiştir. En iyi sonucu elde etmek için farklı müşteri grupları için farklı modeller oluşturulmuş ve puanlar alınmıştır. Analiz sonucunda tekliflere olumlu potansiyele sahip olan müşteriler belirlenmiştir. Böylece, zaman ve maliyet tasarrufu sağlanmıştır. (Taşabat, ve diğ., 2023)

Son ekonomik ve sosyal değişikliklerin ortaya çıkışı, özellikle müşteriler ile perakende mağazaları arasındaki ilişkiyi önemli ölçüde değiştirdi. Geçmişte, perakende sektörü, hizmetlerinden yararlanan müşteriler hakkında detaylı bilgiye sahip olmadan ürünleri pazarlamaya odaklanmıştı. Rakiplerin çoğalması ile, perakende mağazaları müşterilerini tutmaya odaklanmak zorunda kaldı. Bugünkü rekabetçi ortamda başarılı olmak için, perakende mağazalarının müşteri ihtiyaçlarını ve beklentilerini yaratıcı ve yenilikçi bir şekilde karşılamaları gerekmektedir. Genel toplu pazarlama mesajları artık geçerli değildir. Bu çalışma, RFM modeline dayalı yeni bir müşteri sınıflandırma yaklaşımını öne sürmekte ve ayrıca kümeleme ve birliktelik kuralı madenciliği tekniklerini kullanarak müşteri verilerini analiz etme ve müşteri davranışlarını tahmin etme konularıyla ilgilenmektedir (Abirami ve Pattabiraman, 2016).

Yapılan bir çalışmada, Novosibirsk'in endüstriyel kauçuk ürünleri pazarındaki trend verileri analiz edilmiş ve sektörün ana problemleri belirlenmiştir. Endüstriyel kauçuk ürünlerinin üretimi ve satışı ile uğraşan bir şirketin müşteri veritabanına uygulanan modifiye bir MRFM (Modified Recency Frequency Monetary) analizi ile homojen davranışlara sahip müşteri grupları tespit edilmiştir. Her grup ayrıntılı bir şekilde incelenmiştir. Seçilen gruplardaki müşterilerin ihtiyaçlarını en etkili şekilde karşılamak için sadakat programları geliştirilmiştir (Shchekoldin ve Tsoy, 2018).

Yapılan bir çalışmada, tatil rezervasyon sistemine ait veriler kullanılarak mevcut rezervasyon, müşteri verileri işlenmiş ve doğal kümelemeler tespit

edilmiş ve müşterilerin davranışları belirlenmiştir. Bu davranışlara göre hizmetler ve satış stratejileri kişiselleştirilmiştir. K-ortalama uygulamasının ardından, karar ağacı yaklaşımı ile bu mevcut sezgisel bilgileri sağlayan temel özellikler çıkarılmıştır. Bu özelliklerin müşterinin ürün edinme kanalı, belirli ürün tercihleri, rezervasyon dönemleri, mevsimsel tercih vb. olduğu belirlenmiştir. Bu özelliklerin her kümede önemli değişiklikler göstermesi, çözümün genel olarak başarılı olduğunu ve bu özelliklerin başarıyla seçildiğini göstermektedir. Bu çalışma, bu grupların özelliklerine uygun kampanyalar ve ürün paketleri oluşturulmasında önemli bir rol oynamaktadır (Pakyurek ve Sezgin, 2018).

Parikh ve arkadaşı tarafından yapılan bir araştırmada (2020), perakende sektörüne ait online işlemler üzerinde kümeleme algoritmaları ve RFM analizinin müşteri satın alma davranışları için stratejiler sağlamak amacıyla gerçekleştirilebileceğini incelemiştir. Veri seti üzerinde RFM analizi yapmanın yanı sıra, Mean-shift (ortalama kaydırma), Density-Based Spatial Clustering of Applications with Noise (Gürültülü Uygulamaların Yoğunluk Tabanlı Mekansal Kümelenmesi), Agglomerative Clustering (Aglomeratif Kümeleme) ve K-ortalama gibi kümeleme algoritmaları da kullanılmıştır. Bu kümeleme algoritmalarını karşılaştırarak, RFM değerlerine dayalı değerli müşteri grupları bulunmuştur. Bulunan müşteri grupları üzerinde kaybedilecek müşteriyi tutma ve sadık müşteriyi memnun etme gibi etkiler oluşturulabilir (Parikh ve Abdelfattah, 2020).

Bilgi çağında, dijital bilginin işlenmesi ve yayılmasını kapsayan dijital pazarlama, etkili bir pazarlama stratejisi alternatifidir. Endüstriyel sektörde pazarlama, en maliyetli faaliyetlerden biridir ve bu nedenle doğru bir şekilde yürütülmesi gerekmektedir. Purnomo ve arkadaşları tarafından yapılan çalışmada, müşteri satın alma davranışlarına dayanan doğru bir endüstriyel pazarlama stratejisinin oluşturulmasını tartışmaktadır. Müşteri satış verileri, RFM modeli ile ve veri analitiği tekniği ile analiz edilmiştir. Sonuçlar, veri analitiğinin kâr amaçlı olmayan müşterileri dikkate alarak geleneksel RFM'in hassasiyetini artırabileceğini ve gerçek zamanlı veri sağlayabileceğini göstermiştir. Bu araştırma, pazarlamacılar için basit ve yaygın olarak kullanılan bir araç olan geleneksel RFM analizi etkinliğini artırmaya yönelik başka bir bakış açısı sunmaktadır (Purnomo, ve diğ., 2021).

Dursun ve arkadaşı tarafından yapılan arařtırmada, RFM analizi ile otel müşterilerinin profilini ıkarmaya odaklanılmıřtır. Bu alıřmada, Antalya, Trkiye'de faaliyet gsteren  farklı beř yıldızlı otelin gerek CRM verileri kullanılmıřtır. Analiz sonuları, 369 otel müşterisinin sekiz gruba ayrıldıđını gstermiřtir: 'Sadık Müşteriler', 'Sadık Yaz Sezonu Müşterileri', 'Toplu Satın Alma Müşterileri', 'Kıř Sezonu Müşterileri', 'Kayıp Müşteriler', 'Yüksek Potansiyel Müşteriler', 'Yeni Müşteriler' ve 'Kıř Sezonu Yüksek Potansiyelli Müşteriler'. Müşterilerin çođunluđu (%36), daha kısa süreler boyunca konaklayan, diđer gruplara göre daha az harcayan ve otellere yaz sezonunda gelme eđiliminde olan 'Kayıp Müşteriler' segmentinde konumlandırılmıřtır. Sonu olarak, RFM'in müşterileri etkin bir řekilde kümeliyor olmasının, otel üst yöneticilerinin CRM yeteneklerini artırma stratejileri oluřturmasına yol aabileceđini gstermiřtir (Dursun, A. ve Caber, M. 2016)

Son yıllarda, dođrudan pazarlamanın, müşteri veritabanlarını daha etkili bir řekilde segmente etmeyi mümkün kılan veri madenciliđi tekniklerinin kullanılmasıyla daha verimli hale geldiđi gözlemlenmiřtir. Bu konuda mevcut literatürde eřitli analitik teknikler ele alınmaktadır. Örneđin, RFM uzun yıllardır mevcut olan ve poplerliđini basitliđine borlu olan bir tekniktir. Bunun yanında, son yıllarda daha sofistike yöntemlerin geliřtirildiđi görlmüřtür. Ancak, RFM'nin kullanımı hala geniř apta devam etmektedir (McCarty, J., ve Hastak, M. 2007)

Christy tarafından yapılan bir alıřmada (2021) bir řirketin müşterilerinin etkin bir řekilde segmentasyona tabi tutulması ve müşterilerin RFM modeline dayalı olarak benzer davranıř gruplarına ayrılması ele alınmaktadır. Bu alıřma bařlangıta iřlem verilerinde bir RFM analizi gerekleřtirir ve ardından aynıları geleneksel K-ortalama ve bulanık C-ortalama algoritmaları kullanarak kümelemeyi geniřletilmiřtir. K-ortalama algoritmasındaki bařlangı merkezlerini seçmek için yeni bir fikir önerilmiřtir. Elde edilen sonular, iterasyonları, küme sıklıkları ve yürütme süreleri ile birbirleriyle karřılařtırılmıřtır (Christy, ve diđer., 2021).

Yapılan bir arařtırmada dođrudan pazarlamada daha hassas ve bařarılı hedefleme için geleneksel RFM modelini geniřletmeye odaklanılmıřtır. Model, ilk satın alma süresi ve müşterinin devam etmeme olasılıđı gibi iki ek parametre

ile zenginleştirilmiştir. Bu çalışmanın benzersiz bir yönü, bir müşterinin gelecekte bir satın alma yapma olasılığını tahmin etmek ve bir müşterinin gelecekte kaç kez satın alma yapacağını tahmin etmek için olasılık teorisinden Bernoulli dizisinin uygulanmasıdır. Ayrıca, formüldeki bilinmeyen parametreleri tahmin etmek için bir metodoloji önerilmiştir, bu da müşteri hedefleme verimliliğini ve doğruluğunu artırmayı hedeflemektedir (Yeh, I. C., ve diğ., (2009).

Perakende sektöründe bir market zinciri özelinde yapılan çalışmada (2017), müşteri segmentasyonu için yenilikçi bir yaklaşım sunulmuştur ve LRFMP (Length, Recency, Frequency, Monetary, Periodicity - Uzunluk, Yenilik, Sıklık, Parasal Tutar, Periyodiklik) olarak adlandırılan yeni bir model oluşturulmuştur. Market zincirinden elde edilen gerçek hayat verileri kullanılarak, LRFMP modeli ve K-ortalama kümeleme algoritması yardımıyla beş farklı müşteri grubu belirlenmiştir. Bu gruplar "yüksek-katkıda bulunan sadık müşteriler", "düşük-katkıda bulunan sadık müşteriler", "belirsiz müşteriler", "yüksek-harcamalı kaybedilen müşteriler" ve "düşük-harcamalı kaybedilen müşteriler" olarak etiketlenmiştir. Bu araştırma, her müşteri profiline yönelik benzersiz pazarlama stratejilerinin geliştirilmesine yardımcı olmanın yanı sıra, kaynakların verimli bir şekilde tahsis edilmesini sağlar (Peker, S, ve diğ., 2017).

RFM ile yapılan farklı bir çalışmada (2018) Türkiye'nin önde gelen spor malzemeleri perakende zincirlerinden birinin veritabanı kullanılarak, her bir müşteriye bir RFM skoru atanmıştır. Çalışmada kullanılan veri seti, Türkiye'nin en büyük spor malzemeleri perakende zincirlerinden birinin müşteri sadakat kartı veritabanından elde edilmiştir. Bu veri seti, 2016 yılına ait 700032 müşteri kaydını içermektedir. Bu analiz, müşteri davranışlarına uygun stratejiler belirleme ve daha etkin kararlar alma konusunda şirketlere yardımcı olmaktadır (Doğan, O., ve diğ., 2018)

Hu ve Yeh tarafından yapılan çalışmada (2014) perakende sektöründe RFM analizi ile sepet analizinin birlikte ele alındığı görülmektedir. Müşterilerin geçmiş alışveriş davranışlarına dayanarak, RFM analizi değerli bir müşteri grubunu belirleyebilirken, sepet analizi ilginç alışveriş modelleri bulabilmektedir. Bu çalışma, müşteri kimlik bilgisi olmadan, RFM özellikleri göz önüne alınarak model değerlendirmesini doğrudan ölçer. RFM-modeli, sadece sıkça oluşan değil, aynı zamanda son bir alışveriş içeren ve gelirin daha yüksek bir yüzdesini

kapsayan bir model olarak tanımlanır. Çalışma ayrıca tüm işlem veritabanını sıkıştırmak ve depolamak için bir ağaç yapısı olan bir RFM-model-ağacı önerir ve bir RFM-model-ağacındaki tüm RFM-modelleri keşfetmek için model büyüme tabanlı bir algoritma olan RFMP-büyüme geliştirir. Deneysel sonuçlar, önerilen yaklaşımın etkili olduğunu ve RFM-müşteri-modellerinin büyük bir kısmını başarılı bir şekilde keşfedebildiğini göstermektedir. (Hu, Y., ve diğ., 2014)

Hedef odaklı pazarlama stratejileri, müşteri satın alma döngüsündeki heterojenliği anlamak için pazar segmentasyonunu kullanır. Ancak, bu genellikle müşteri davranışının zamanla nasıl evrimleştiğini göz ardı eder ve bu da perakendecilerin kar getirmeyen müşterilere odaklanmasına neden olur. Yoseph ve arkadaşı tarafından yapılan çalışmada (2018), Orta büyüklükteki bir Kuveyt giyim ve moda aksesuarı perakendecisinin müşterilerini segmente etmek için RFM ile Müşteri Yaşam Boyu Değer modeli bir arada kullanılmaktadır. Modifiye edilmiş bir regresyon algoritması, satış noktası verilerinden bilgi edinerek müşteri satın alma eğrisini incelemekte ve böylece perakendecinin bilinçli kararlar vermesine yardımcı olmaktadır. Kümeleme, K-ortalama ve Beklenti Maksimizasyonu ile gerçekleştirilir. Küme kalitesi analizi, K-ortalamanın Beklenti Maksimizasyonu'na göre daha iyi performans göstererek ilgili pazar segmentlerini keşfetme ve uygun pazarlama stratejileri önerme konusunda daha üstün olduğunu ortaya koyar (F. Yoseph ve M. Heikkila, 2018)

Perakende sektöründe kampanya yönetimi üzerine Martinez ve arkadaşları tarafından yapılan bir vaka çalışmasında (2019) geleneksel RFM modeline ve bir müşteri veritabanına uygulanan 2-tuple RFM modeline dayalı bir müşteri segmentasyonunun ne kadar kolay, doğru ve açıklanabilir olabileceği konusunda gerçek bir örnek gösterilmektedir. Bu, geleneksel model yerine 2-tuple modelin uygulanmasının faydalarını daha iyi anlamamızı sağlayacaktır. 2-tuple modelin üzerine K-ortalama kümelemesi uygulayarak, segmentlerin iş açısından büyük bir uygulanabilirliğe sahip olduğu görülmektedir. Tanımlayıcı değişkenler kullanarak, küme tanımını açıklığa kavuşturacağız ve model bize müşterilerin nasıl davrandığı hakkında son derece net bir fikir verecektir. Bu örneği geliştirmenin ana hedefi, bir doğrudan kampanya iletişimi için en iyi hedefi tanımlamaktır. Bu analiz için kullanılan veriler, bir dünya çapında ev mobilyası,

İskandinav perakendecisine ait olup, üyelerin tarihsel satın alma bilgilerini bize veren sadakat programıyla ilgilidir" (González Martínez, ve diğ., 2019)

Wan ve arkadaşları tarafından yapılan çalışmada RFM Analizi'nin müşteri segmentasyonu ve müşteri yaşam değeri tahmini gibi alanlarda etkili bir araç olduğunu belirtmiştir. Çalışmalarında, RFM Analizi'nin, müşteri davranışlarını daha iyi anlamak ve müşteri ilişkileri yönetimini geliştirmek için kullanılabileceğini göstermişlerdir. Ayrıca, RFM Analizi'nin, müşteri sadakatini ve satışları artırmak için pazarlama stratejilerini geliştirmeye yardımcı olabileceğini belirtmişlerdir. Sonuçlarına göre, RFM Analizi'nin, müşteri segmentasyonu ve hedefleme stratejilerinin geliştirilmesinde önemli bir rol oynayabileceği sonucuna varmışlardır. (Wan, S., ve diğ., 2022)

Aşağıda RFM özelinde yapılan literatür incelemesine yönelik bir özet tablo verilmiştir.

Yazar ve Araştırma	Yöntem	Özet
Cheng, C. H., & Chen, Y. S. (2009). Classifying the segmentation of customer value via RFM model and RS theory. Expert Systems with Applications,	RFM, K-Ortaama	RFM bileşenleri ve K-means kümeleme algoritması kullanılarak CRM için üstün bir sınıflandırma kural listesi oluşturmayı önerir. Bu model, veri madenciliği teknolojilerinin olumsuz etkilerini azaltır.
Chen, Y. L., Kuo, M. H., Wu, S. Y., & Tang, K. (2009). Discovering recency, frequency, and monetary (RFM) sequential patterns from customers' purchasing data. Electronic Commerce Research and Applications,	RFM, LTV	Müşteri segmentasyonu ve pazarlama kampanyası arasındaki bağlantı eksikliğini ele alır. RFM modeli ve Müşteri Yaşam Boyu Değer (LTV) modeli kullanarak müşteri segmentasyonu ve hedeflemeyi birleştirir.
Wei, Jo-Ting & Lin, Shih-Yen & Wu, Hsin-Hung. (2010). A review of the application of RFM model. African Journal of Business Management December Special	RFM	RFM modelinin, kâr amacı güden ve gütmeyen kuruluşlar için değerli müşteri belirlemede ve pazarlama stratejileri geliştirmede etkili olduğunu gösterir.
Khajvand, M., Zolfaghar, K., Ashoori, S., & Alizadeh, S. (2011). Estimating customer lifetime value based on RFM analysis of customer purchase behavior: Case study.	RFM, CLV	Müşteri değeri konsepti üzerinden CRM oluşturma amacını ele alır. Genişletilmiş RFM analizi ve 'Count Item' parametresi kullanılarak müşteri segmentasyonu yapılır.

Kim, D., Lee, J. Y., Ahn, S., Moon, Y., & Kwon, O. J. (2012). RFM analysis for detecting future core technology.	RFM	RFM analizinin, gelecekteki çekirdek teknolojileri tahmin etmek için kullanılabileceğini belirtir. ABD patent verileri temel alınarak bu analiz gerçekleştirilmiştir.
You, Z., Si, Y. W., Zhang, D., Zeng, X., Leung, S. C., & Li, T. (2015). A decisionmaking framework for precision marketing	RFM, CHAID	Hassas pazarlama için yeni bir karar verme çerçevesi sunar. RFM modeli, CHAID karar ağaçları ve Pareto değerleri kullanılarak müşteri segmentasyonu yapılır.
Nikumanesh, E., & Albadvi, A. (2014). Customer's life-time value using the RFM model in the banking industry	RFM	RFM modeli kullanarak İran Tarım Bankası'nın müşteri veritabanını inceleyerek müşteri segmentasyonu yapar. Sonuçlar, normal müşterilerin en düşük karlılığa sahip olduğunu gösterir.
Coussement, K., van den Bossche, F. A., & de Bock, K. W. (2014). Data accuracy's impact on segmentation performance: Benchmarking RFM analysis, logistic regression, and decision trees.	RFM	Veri kalitesinin segmentasyon tekniklerinin performansı üzerindeki etkisini inceler. Karar ağaçlarının, RFM analizi ve lojistik regresyondan daha üstün olduğu sonucuna varır.
Sarvari, P. A., Ustundag, A., & Takci, H. (2016). Performance evaluation of different customer segmentation approaches based on RFM and demographics analysis.	RFM, WRFM	Müşteri segmentasyonuna en iyi yaklaşımın ne olduğunu belirlemek için RFM ve demografik faktörlere dayalı segmentasyon yaklaşımlarını değerlendirir. Demografik verilerin WRFM ile birleştirilmesinin önemini vurgular.
Dursun, A., & Caber, M. (2016). Using data mining techniques for profiling profitable hotel customers: An application of RFM analysis. Tourism management perspectives,	RFM	Antalya'daki beş yıldızlı otellerin CRM verileri kullanılarak RFM analizi ile otel müşteri segmentasyonu yapılmıştır. En büyük segment 'Kayıp Müşteriler' olarak belirlenmiştir.
Abirami, M., & Pattabiraman, V. (2016). Data Mining Approach for Intelligent Customer Behavior Analysis for a Retail Store.	RFM	Perakende sektöründe müşteri ihtiyaçlarına yaratıcı bir şekilde yanıt vermek için RFM modeline dayalı bir müşteri sınıflandırma yaklaşımı sunulmuştur.
Shchekoldin, V. Y., & Tsoy, M. Y. (2018). The Application of Modified RFM Analysis to Increase the Loyalty of Consumers of Industrial Rubber Articles.	MRFM	Endüstriyel kauçuk ürünleri pazarında MRFM-analizi kullanılarak müşteri segmentasyonu yapılmış ve sadakat programları geliştirilmiştir.

Pakyurek, M., Sezgin, M. S., Kestepe, S., Bora, B., Duzagac, R., & Yildiz, O. T. (2018). Customer clustering using RFM analysis.	RFM, K-ortalama	Tatil rezervasyon verileri üzerinde RFM ve K-means kullanılarak müşteri segmentasyonu yapılmıştır.
Parikh, Y., & Abdelfattah, E. (2020). Clustering Algorithms and RFM Analysis Performed on Retail Transactions.	RFM, K-ortalama ,DBSCAN	Perakende sektöründe RFM ve çeşitli kümeleme algoritmaları kullanılarak müşteri segmentasyonu gerçekleştirilmiştir.
Purnomo, Muhammad & Anugerah, Adhe & Azzam, Abdullah & Khasanah, Annisa & Alfareza, M. (2021). RFM-based Customers Clustering for Precise Industrial Marketing Strategy Formulation.	RFM	Endüstriyel pazarlamada RFM analizi ile müşteri segmentasyonu yapılarak pazarlama stratejilerinin etkinliği artırılmıştır.
McCarty, J., & Hastak, M. (2007). Segmentation approaches in data-mining: A comparison of RFM, CHAID, and logistic regression.	RFM, CHAID	RFM, CHAID ve lojistik regresyon gibi veri madenciliği teknikleri doğrudan pazarlama için karşılaştırılmıştır.
Doğan, O., Ayçin, E., & Bulut, Z. (2018). Customer segmentation by using RFM model and clustering methods: a case study in retail industry.	RFM	Türkiye'nin spor malzemeleri perakende sektöründe RFM kullanılarak müşteri segmentasyonu yapılmıştır.
Hu, Y. H. & Yeh, T. W. (2014). Discovering valuable frequent patterns based on RFM analysis without customer identification information	RFM	RFM analizi ve sepet analizi bir araya getirilerek müşteri alışveriş modelleri keşfedilmiştir.
F. Yoseph and M. Heikkila,(2018) "Segmenting Retail Customers with an Enhanced RFM and a Hybrid Regression/Clustering Method,"	RFM	RFM ve Müşteri Yaşam Boyu Değer modeli bir araya getirilerek müşteri segmentasyonu yapılmıştır.
Martínez, Carrasco, Madariaga, Gallego, Viedma, (2019) "A comparison between Fuzzy Linguistic RFM Model and traditional RFM model applied to Campaign Management.	RFM	Geleneksel RFM modeli ile 2-tuple RFM modeli karşılaştırılarak müşteri segmentasyonu gerçekleştirilmiştir.
F. A. Bachtiar, "Customer Segmentation Using Two-Step Mining Method Based on RFM Model,"	RFM	RFM analizi kullanılarak müşteri segmentasyonu yapılmış ve segmentasyon sonuçlarına göre pazarlama stratejileri geliştirilmiştir.

Köse, U., & Arslan, A. (2020). A Novel Customer Segmentation Approach Based on RFM and Clustering: A Case Study in the Retail Industry.	RFM	RFM modeli kullanılarak müşteri segmentasyonu yapılmış ve segmentasyon sonuçlarına göre pazarlama stratejileri önerilmiştir.
Wan, S., Chen, J., Qi, Z., Gan, W., & Tang, L. (2022, April). Fast RFM model for customer segmentation.	RFM	RFM Analizi'nin müşteri segmentasyonu için etkili bir araç olduğu belirtilmiştir.
Miglautsch, J. R. (2000). Thoughts on RFM scoring. Journal of Database Marketing & Customer Strategy Management,	RFM	RFM skorlamasının müşteri segmentasyonunda etkili bir yöntem olduğu vurgulanmıştır.
Erpolat Taşabat, S., & Akca, E. (2020). Recycling Project with RFM Analysis in Industrial Material Sector. Sigma Journal of Engineering and Natural Sciences,	RFM	RFM analizi kullanılarak geri dönüşüm projeleri için müşteri segmentasyonu gerçekleştirilmiştir.
Christy, A.J. (2018). RFM ranking – An effective approach to customer segmentation.	RFM	RFM analizi ve çeşitli kümeleme yöntemleri kullanılarak müşteri segmentasyonu yapılmıştır.
Taşabat, S. E., Özçay, T., Sertbaş, S., & Akca, E. (2023). A New RFM Model Approach: RFMS. In Industry 4.0 and the Digital Transformation of International Business	RFM, RFMS	Klasik RFM modeline ek olarak ekonomik değişkenin eklenmesiyle RFMS modeli önerilmiştir.
Barus, O. P., Nathasya, C., & Pangaribuan, J. J. (2023). The Implementation of RFM Analysis to Customer Profiling Using K-Means Clustering	RFM, K-Ortalama	RFM analizi ve K-Means kümeleme yöntemi kullanılarak müşteri segmentasyonu gerçekleştirilmiştir.

Şekil 1. RFM Literatür Taraması Özet Tablo

III. METODOLOJİ

A. Makine Öğrenmesi

Makine öğrenmesi, bilgisayarların belirli bir programlama olmadan öğrenme ve karar verme yeteneği kazanmasını sağlayan bir yapay zeka alt dalıdır. Bu, bilgisayarların verilerden öğrenmesi ve bu öğrenimleri yeni verilerle karşılaştıklarında tahminlerde bulunmak veya kararlar vermek için kullanması anlamına gelir. Makine öğrenmesi, performansı artırmak veya doğru tahminler oluşturmak için önceden bilgiye dayanan hesaplamalı algoritmaları ifade eder. Bu durumda, deneyim, genellikle toplanan veri şeklinde ve modellemeye hazır olan öğrenenin önceki bilgisini ifade eder. Bu bilgi, etiketli ikili veya daha fazla eğitim setleri şeklinde veya çevreyle etkileşim yoluyla toplanan diğer türden veriler şeklinde olabilir. Her durumda, kalitesi ve büyüklüğü, öğrenenin tahminlerinin başarısı için kritiktir. Etkili ve doğru tahmin algoritmaları tasarlamak, makine öğrenmesinin temelini oluşturur. Ancak, makine öğrenmesinde, bir algoritmanın bir dizi fikri öğrenmek için gereken örnek büyüklüğünü belirlemek için örnek karmaşıklığı kavramına da ihtiyaç duyulmaktadır. Fikir sınıflarının karmaşıklığı ve eğitim örneğinin büyüklüğü, bir algoritma için teorik öğrenme garantilerini belirler. Bir öğrenme algoritmasının etkinliği, kullanılan verilere dayandığı için veri analizi ve istatistiklerle bağlantılıdır. Genel olarak, öğrenme teknikleri, çekirdek bilgisayar bilimleri kavramlarını istatistik, olasılık ve optimizasyon kavramlarıyla birleştiren veriye dayalı yöntemlerdir (Trappenberg, 2019).

Makine öğrenmesi modellemesi yöntemleri - model seçimi, çapraz doğrulama, örnekleme vb. dahil olmak üzere çeşitli değişken setlerini kullanarak modeli eğitme ve seçilen modellerden tahminler yapmayı kapsar (Matuszelański, K., ve Kopczewska, K. 2022).

Makine öğrenmesi görevlerinin çoğu, denetimli veya denetimsiz olmak üzere iki kategoriye ayrılır. Denetimli öğrenme, etiketli veri kümelerini kullanır

ve bir tür makine öğrenmesidir. Bu veri kümeleri, bilgisayarların verileri nasıl tanıyacağını ve sonuçları nasıl tahmin edeceğini öğrenmesi için kullanılır. Etiketli girişler ve çıktılar kullanılarak model, doğruluk için test edilebilir ve zamanla öğrenebilir. Denetimli öğrenmede iki tür problem vardır: sınıflandırma ve regresyon (James, 2013).

1. Denetimli Öğrenme

Denetimli öğrenme, bir algoritmanın giriş verileri (özellikler) ve çıktı verileri (etiketler) arasındaki ilişkiyi öğrenmeye çalıştığı bir makine öğrenmesi türüdür. Bu tür öğrenme, genellikle bir eğitim veri seti kullanılarak gerçekleştirilir. Eğitim veri seti, her bir girişin beklenen çıktısını (etiketi) içerir. Algoritma, giriş verileri ve etiketler arasındaki ilişkiyi modellemeye çalışır.

Denetimli öğrenme, genellikle iki ana kategoriye ayrılır: sınıflandırma ve regresyon.

1. Sınıflandırma: Sınıflandırma, bir girişin belirli bir sınıfa ait olup olmadığını tahmin etmeye çalışır. Bu, genellikle iki sınıf (ikili sınıflandırma) veya daha fazla sınıf (çok sınıflı sınıflandırma) arasında bir ayırım yapar. Örneğin, bir e-postanın spam olup olmadığını belirlemek için bir sınıflandırma algoritması kullanılabilir.
2. Regresyon: Regresyon, bir girişin belirli bir sayısal değeri tahmin etmeye çalışır. Bu, genellikle sürekli bir çıktı değeri tahmin etmek için kullanılır. Örneğin, bir evin fiyatını tahmin etmek için bir regresyon algoritması kullanılabilir.

Denetimli algoritmalar, x ve y veri kümesi ile, bazı girdileri bazı sonuçlarla ilişkilendirmeyi öğrenir. Eğitim seti hedefleri otomatik olarak elde edilmiş olsa bile, y çıktıları otomatik olarak elde edilemez ve bir insan "supervizör" tarafından sağlanmalıdır (Goodfellow, 2016). "Denetim" terimi, insan katılımını ifade etmez; bunun yerine hedef değerlerin, öğrenci için denetleyici bir rol oynadığını, öğrenilmesi gereken görevi belirttiğini ifade eder. Öğrenme algoritması, istenen çıktıyı üreten özellik değerlerinin kombinasyonunu bulmaya çalışır (Lantz, 2013).

2. Denetimsiz Öğrenme

Denetimsiz öğrenme, makine öğrenmesinin en önemli kategorilerinden biridir. Denetimsiz öğrenme algoritmaları, etiketlenmemiş veriler üzerinde çalışır. Yani, girdi verileri üzerinde herhangi bir hedef veya sonuç değişkeni yoktur. Bu tür algoritmalar, veri setindeki gizli yapıları veya düzenleri bulmaya çalışır. Denetimsiz öğrenme, genellikle büyük veri setlerindeki gizli kalıpları ve yapıları keşfetmek için kullanılır. Ancak, denetimsiz öğrenme algoritmalarının sonuçlarını doğrulamak veya değerlendirmek genellikle zordur çünkü doğru cevaplar genellikle bilinmemektedir.

Denetimsiz öğrenme genellikle etiketsiz veri setinden değerli bilgiler elde etmeye odaklanır. Verinin en iyi temsili, klasik bir gözetimsiz öğrenme görevidir. "En iyi" terimi çeşitli şeyleri ifade eder, ancak genel olarak ana hedef, x hakkında mümkün olduğunca çok bilgiyi korurken, temsili daha basit veya x 'in kendisinden daha erişilebilir kılan bazı doğruluk metrikleri veya kısıtlamalara uygun bir temsildir (Goodfellow, 2016).

B. Kümeleme

Kümeleme, etiketsiz bir veri kümesini kullanarak kümelere nasıl etiket atayacağımızı öğrenmeye çalıştığımız karmaşık bir süreçtir. Veri kümesi tamamen etiketsiz olduğundan, öğrenilen modelin en iyi model olup olmadığını belirlemek, denetimli öğrenme durumlarına kıyasla çok daha zordur. Maalesef, hangi kümeleme algoritmasının sizin veri kümeniz için en uygun olduğunu belirlemek oldukça güçtür çünkü kullanabileceğiniz çok sayıda algoritma bulunmaktadır. Genelde, bir algoritmanın performansı, veri kümesinin çıkarıldığı olasılık dağılımının bilinmeyen parametrelerine bağlıdır. Bu durum, kümeleme algoritmalarının etkinliğini değerlendirmeyi zorlaştırır (Burkov, 2019).

1. K-Ortalama

K-ortalama algoritması, yaygın bir kümeleme yöntemidir ve en az iki parametre gerektirir: küme sayısı (k) ve uzaklık ölçüm metodu. Algoritma, veri noktalarının küme merkezlerine olan uzaklıklarını hesaplar ve kümeleri belirler. Bu işlem, belirlenen durma kriterlerinden biri karşılanana kadar tekrarlanır. Durma kriterleri genellikle küme merkezlerinde bir değişiklik olmaması veya

belirli bir iterasyon sayısına ulaşılmasıdır. K-ortalama, hızlı ve etkili bir yöntem olmasına rağmen, başlangıç merkezlerinin rastgele seçilmesi ve küme sayısının doğru belirlenmesi gibi faktörler sonuçları etkileyebilir. Bu nedenle, birden fazla çalışma yapılması ve en iyi sonucun seçilmesi gerekmektedir (James, G., ve diğ., 2013).

$$E = \sum_{i=1}^k \sum_{p \in c} d(M_i, x)^2$$

Formu 1

K-ortalama algoritması, geniş çapta kullanılan bir kümeleme yöntemidir. Bu algoritmanın çeşitli varyasyonları bulunsa da temel prensipler ve algoritmanın işleyiş süreci genellikle aynıdır. K-ortalama algoritması, her çalıştırılmasında belirli parametreler talep eder. En basit varyasyonu bile en az iki parametre gerektirir: küme sayısı (k) ve uzaklık ölçüm metodu. Algoritma, kümeye dahil olan veri noktalarındaki değişkenlerin ortalamalarını hesaplayarak küme merkezlerini belirler. Ardından, veri noktalarının seçilen uzaklık metoduna göre küme merkezlerine olan uzaklıkları hesaplanır ve kümeler belirlenir. Bu işlem, belirlenen durma kriterlerinden biri karşılanana kadar tekrarlanır.

Algoritma, belirlenen küme sayısı (K) ve değişken sayısı (N) büyüklüğündeki (KxN) bir matrisin rastgele doldurulmasıyla başlar. Rastgele küme merkezi seçimi, algoritmanın her çalıştırılmasında farklı sonuçlar elde edilmesine neden olur. Özellikle dağılımı zor olan değişkenlerde, algoritmanın her bir yapılandırması için (parametre seçimi) birden fazla çalıştırma yapılması gerekebilir. İkinci aşamada, veri noktalarının küme merkezlerine olan uzaklıkları hesaplanır. Uzaklık hesaplamalarında vektörel ve skaler olacak şekilde farklı varyasyonlar kullanılabilir. Bu çalışmada, uzaklık hesaplama metodu olarak Öklidyen uzaklık seçilmiştir. Uzaklık hesaplamaları sonucunda, her veri noktası en yakın olduğu kümeye atanır ve böylece veri noktalarının ilk kümeleri belirlenir.

Algoritmanın üçüncü aşamasında, yapılan atamalar baz alınarak küme merkezleri yeniden hesaplanır. Küme merkezleri, kümeye dahil olan veri noktalarının her bir değişken için ayrı ayrı ortalamaları alınarak belirlenir. Küme merkezi sayısı kadar alınan ortalama değerler, algoritmanın adını oluşturur. Yeni küme merkezleri belirlendikten sonra, ikinci aşamada olduğu gibi, veri noktaları

yeni oluşan küme merkezlerine olan uzaklıklarına göre yeniden kümelendir. Bu işlem, durma kriterlerinden herhangi biri karşılanana kadar tekrarlanır. K-ortalama algoritmasında genellikle iki farklı durma kriteri kullanılır. Bu kriterlerden en yaygın kullanılanı, küme merkezlerinin yeniden hesaplanması ve noktaların kümeleneceği sonucunda bir önceki iterasyonla veya belirlenen iterasyon boyunca değişiklik olmaması durumunda algoritmanın durdurulmasıdır. K-ortalama algoritması, hesaplama karmaşıklığı teorisine göre NP karmaşıklık sınıfında bir problem olduğu için, veri setinin büyüklüğü arttıkça birinci durma kriterinin sağlanması büyük bir işlem gücü gerektirir. Bu nedenle, ikinci en yaygın kullanılan durma kriteri, üçüncü aşamadan itibaren başlayan iterasyon sayısının sınırlanmasıdır.

K-ortalama algoritması genellikle hızlı ve etkili bir kümeleme yöntemi olmasına rağmen, başlangıç merkezlerinin rastgele seçilmesi sonuçları etkileyebilir. Bu nedenle, farklı başlangıç noktalarıyla birden fazla çalışma yapılması ve en iyi sonucun seçilmesi önerilir. Ayrıca, küme sayısının doğru bir şekilde seçilmesi ve veri setine uygun bir değer belirlenmesi de önemlidir. Bu faktörlerin dikkate alınması, K-ortalama algoritmasının daha etkili ve doğru sonuçlar vermesini sağlar.

2. Bulanık C-Ortalama

Kümeleme, gözetimsiz makine öğreniminin öne çıkan bir görevidir ve benzer nesnelere gruplama süreci olarak tanımlanır. Genel olarak, kümeleme, benzerliklere dayalı olarak veri noktaları veya örneklerin doğal gruplamalarını keşfetmeyi amaçlar. Bu bağlamda, birçok kümeleme yöntemi literatürde geliştirilmiştir ve kümelemenin bilim ve mühendislikteki farklı uygulama alanlarında, veri madenciliği, makine öğrenimi, arıza teşhisi, desen tanıma ve arıza teşhisi gibi alanlarda kritik bir rol oynadığı görülmektedir. Bulanık C-ortalama, Bezdek tarafından önerilen ve en yaygın kullanılan bulanık kümeleme yöntemlerinden biridir. Bulanık C-ortalama kullanımının ana avantajı, veri kümesi özelliklerinden dolayı kümelerin doğada örtüşüyor olması durumunda, bulanık C-ortalama genellikle daha yüksek seviye bilgi işleme sonucu daha başarılı olmasıdır. (Bezdek, J. C., ve diğ., 1984)

Bulanık C-ortalama, veri vektörlerini küme merkezlerine olan uzaklıklarına dayanarak belirli bir nesnenin bir kümeyle olan üyeliğini belirlemek için kullanılan bir yöntemdir. Bulanık C-ortalama algoritması, aşağıdaki amaç fonksiyonunu minimize ederek türetilir. Bu, tüm nesnelerin üyeliklerini hesapladıktan sonra kümelerin yeni prototiplerini hesaplamak için kullanılır. Süreç, prototipler stabil hale geldiğinde durur. Yani, önceki yinelemeden gelen prototipler, genellikle bir hata eşiğinden daha az olan mevcut yinelemede üretilenlere yakındır. Bulanık C-ortalama, veri vektörlerini küme merkezlerine olan uzaklıklarına dayanarak belirli bir nesnenin bir kümeyle olan üyeliğini belirlemek için kullanılan bir yöntemdir. Bulanık C-ortalama'nın en popüler bulanık kümeleme algoritması olduğu ve Bezdek tarafından tanıtıldığı ve şu anda yaygın olarak kullanıldığı belirtilmiştir. FCM etkili bir algoritmadır, ancak merkez noktalarındaki rastgele seçim, yinelemeli sürecin eyer noktalarına veya yerel en iyi çözüme kolayca düşmesine neden olmaktadır. (Zhang, J., ve Shen, L. 2014)

Bulanık C-ortalama algoritması, genellikle bulanık kümeleme problemlerini çözmek için kullanılan bir yöntemdir. Bu algoritma, K-ortalama algoritmasının bir varyasyonu olup, veri noktalarının birden fazla kümeye ait olabileceğini kabul eder. Bulanık C-ortalama algoritması, en az iki parametre gerektirir: küme sayısı (c) ve bulanıklık derecesi (m). Algoritma, her bir veri noktasının her bir kümeye aidiyet derecesini hesaplar ve bu aidiyet derecelerinin ortalamasını alarak küme merkezlerini belirler. Ardından, veri noktalarının küme merkezlerine olan uzaklıkları hesaplanır ve aidiyet dereceleri güncellenir. Bu işlem, belirlenen durma kriterlerinden biri karşılanana kadar tekrarlanır. Bulanık C-ortalama algoritmasında genellikle iki farklı durma kriteri kullanılır. Birincisi, küme merkezlerinin yeniden hesaplanması ve aidiyet derecelerinin güncellenmesi sonucunda bir önceki iterasyonla veya belirlenen iterasyon boyunca değişiklik olmaması durumunda algoritmanın durdurulmasıdır. İkincisi, belirli bir iterasyon sayısına ulaşıldığında algoritmanın durdurulmasıdır. Bulanık C-ortalama algoritması, genellikle etkili bir kümeleme yöntemi olmasına rağmen, başlangıç merkezlerinin ve bulanıklık derecesinin seçimi sonuçları etkileyebilir. Bu nedenle, farklı başlangıç noktaları ve bulanıklık dereceleri ile birden fazla

çalışma yapılması ve en iyi sonucun seçilmesi gerekmektedir (Baykasoğlu, A., ve diğ., 2018).

$$J_m = \sum_{i=1}^N \sum_{j=1}^c u_{ij}^m \|x_i - c_j\|^2, 1 < m < \infty$$

Formu 2

C. Değerlendirme Yöntemleri

Hem K-ortalama hem de bulanık C-ortalama algoritmalarında olduğu gibi en iyi sonucu belirlemek için birden çok çalışma yapılması ve yapılan bu çalışmalarının sonucunun sistematik olarak değerlendirilmesi gerekmektedir. Bu nedenle Silhouette indeks ile kümelerin performansı değerlendirilecektir.

1. Silhouette indeksi

Silhouette indeks, kümeleme performansını ölçmek için kullanılan bir metriktir. Hem K-ortalama hem de bulanık C-ortalama gibi kümeleme algoritmalarının sonuçlarını değerlendirmek için yaygın olarak kullanılır. Silhouette indeks, bir veri noktasının kendi kümeleri içindeki diğer noktalara olan uzaklığına (bireysel uyum) ve en yakın diğer kümeye olan uzaklığına (bireysel yabancılık) dayanır. Silhouette Index, her bir veri noktası için bir değer hesaplar. Bu değer, -1 ile 1 arasında bir aralıktadır. Bir veri noktasının Silhouette indeks değeri 1'e yakınsa, bu, veri noktasının kendi kümesine iyi bir şekilde uyduğunu ve diğer kümelerden uzak olduğunu gösterir. Eğer değer -1'e yakınsa, bu, veri noktasının kendi kümesine iyi bir şekilde uymadığını ve belki de başka bir kümeye ait olması gerektiğini gösterir. Eğer değer 0'a yakınsa, bu, veri noktasının kümeler arasında sınırda olduğunu ve hangi kümeye ait olduğunun belirsiz olduğunu gösterir (Starczewski, A., ve Krzyżak, A. 2015).

$$S(x) = \frac{b(x) - a(x)}{\max(a(x), b(x))}$$

Formu 3

D. RFM Analizi

RFM Analizi, müşteri segmentasyonu ve doğrudan pazarlama için kullanılan bir tekniktir. RFM, bir müşterinin ne zaman (Recency), ne sıklıkla (Frequency) ve ne kadar para harcadığını (Monetary) ölçer. Bu üç ölçüm, bir

müşterinin bir işletme için ne kadar değerli olduğunu belirlemek için birlikte kullanılır. RFM analizi, müşterileri bu üç ölçüme göre sıralar ve her birini belirli bir puanla değerlendirir. Bu puanlar daha sonra bir müşteri değerlendirme skoru oluşturmak için birleştirilir. Bu skor, müşterileri segmentlere ayırmak ve her bir segmente özelleştirilmiş pazarlama stratejileri uygulamak için kullanılır. RFM, büyük miktarda veriden önemli müşterileri ayırmak için üç özelliği kullanan bir modeldir (Sarvari, 2016)

Müşteriler, son zamanlarda alışveriş yaptıkları bir şirketi ve markasını gelecekteki alışverişler için daha kolay hatırlarlar. Son zamanlarda bir şirketten ödeme yapmış olan tüketicilerin, aylar hatta daha uzun süre boyunca şirketten alışveriş yapmamış olan müşterilere kıyasla gelecekte başka bir alışveriş yapma olasılıkları daha yüksektir. Bu tür bilgiler, son müşterileri geri dönmeye ve daha fazla para harcamaya teşvik etmek için kullanılabilir. Eski müşterilerin izini kaybetmemek için, onlara son işleminden bu yana bir süre geçtiğini hatırlatırken aynı zamanda başka bir alışveriş yapmaları için bir teşvik sağlamak üzere pazarlama çabaları yapılmalıdır. Miglautsch'a göre (2000), müşteri kazanma ve iletişimde bilimsel yöntem, doğrudan pazarlamanın kalbidir. Ana nokta, müşterinin önceki alışveriş geçmişine dayanarak bir takip çağrısına layık olup olmadığıdır. Bu soru, doğrudan posta, kataloglar, telefon, saha ve internet dahil olmak üzere her tür iletişim için geçerlidir. Bu seçimi yapma prosedürüne müşteri segmentasyonu denir. RFM skorlamasının amacı, gelecekteki davranışı öngörmektir. Müşteri verilerini istatistiksel dile getirmek çok önemlidir (Miglautsch 2000). Segmentasyona başlamak için öncelikle yakınlık, sıklık ve parasal değerlerin alınması gerekmektedir. Bundan sonra müşteriler azalan sırayla sıralanır ve bu tür bir sıralama, değerlerin ölçümünü sağlar. 1'den 4'e kadar her değer dağılımı puanlanır ve bu, istatistiklere dayanır.

RFM modeli, genellikle doğrudan gelecekteki sonuçları tahmin etmek yerine çıkarımsal bilgiler sağlamak için kullanılır (Yang, 2004).

Yang'a göre RFM Analizi'nin güçlü yönleri aşağıdaki gibidir:

1. Basitlik ve Eyleme Geçirilebilirlik: RFM, müşteri davranışlarını anlamak ve gelecekteki eylemleri tahmin etmek için basit ve eyleme geçirilebilir bir

yöntemdir. Bu, RFM'nin genellikle pazarlama ve satış ekipleri tarafından kolayca anlaşılabilen bir model olmasını sağlar (Yang, 2004).

2. Esneklik: RFM kodlaması oldukça keyfidir ve hem artıları hem de eksileri vardır. İyi bir yanı, ampirik yöntemin genellikle esnek bir kodlama sistemi ile her zaman çalışabilir olmasıdır. Bu, RFM'nin terk edilmemesi gerektiğini, ancak pazar çabalarının gücünü artırmak için önemli bir öngörücü olarak kullanılması gerektiğini gösterir (Yang, 2004).
3. Gelecekteki Yanıt Olasılığını Çıkarabilme: RFM prensibine uygun olarak, V (Value - Değer) geçmiş değeri temsil eder ve gelecekteki yanıt olasılığını çıkarır. Bu, RFM'nin müşteri davranışlarını ve gelecekteki eylemleri tahmin etme yeteneğini güçlendirir (Yang, 2004).

1. Recency

RFM Analizi'nde "Recency" (R), bir müşterinin en son satın alma işleminin ne kadar "yakın" olduğunu ifade eder. Bu, genellikle bir müşterinin en son satın alma işleminden itibaren geçen zaman aralığı olarak ölçülür. Recency, müşterinin işletmeyle olan etkileşiminin tazeliğini gösterir ve genellikle müşterinin gelecekteki satın alma olasılığını tahmin etmek için kullanılır.

Müşteri davranışlarını anlamak için yakın zamanda ödeme yapmanın önemi göz önüne alındığında, yakınlık ön plana çıkar. Sektörde, insanlar iletişim kararlarını sadece son iletişimlerinin ne kadar yakın olduğuna dayanarak yaparlar. Kısa bir süre önce alışveriş yapan müşteriler aktif ve geçerli kabul edilir. Bu müşteriler ilk 12 ay boyunca oldukça sık pazarlama faaliyetlerine maruz kalırlar, daha sonra 36-48 aya kadar daha az sıklıkla erişim sağlanır. Daha sonrasında ise geri dönüş alınamayan bir şekilde alışveriş yapmayan müşteriler ise "churn" yani kaybedilmiş olarak adlandırılabilir (Birant, 2011).

2. Frequency

RFM Analizi'ndeki "F" yani "Frequency" (Sıklık), bir müşterinin belirli bir zaman diliminde ne kadar sık alışveriş yaptığını ifade eder. Bu, genellikle müşterinin bir işletmeyle olan etkileşim sıklığını gösterir ve müşterinin işletmeye olan bağlılığının bir göstergesi olarak kabul edilir. Daha sık alışveriş yapan müşteriler genellikle daha değerli kabul edilir çünkü bu, onların işletmeyle daha

güçlü bir ilişkisi olduğunu ve gelecekte de alışveriş yapma olasılıklarının daha yüksek olduğunu gösterir.

Müşterinin ne kadar sıklıkla işlem yaptığını etkileyebilecek faktörler arasında malın türü, satın alma fiyatı ve yeniden tedarik etme veya değiştirme gereksinimi bulunabilir. Bir müşterinin alışveriş döngüsü öngörülebilirse, örneğin ekstra bir satın alma yapması gerektiğinde, pazarlama çabaları yapılabilir. Bununla birlikte temel ürünlerinin tükendiğinde mağazayı ziyaret etmelerini hatırlatmaya yönlendirilebilir. Sıklık skoru söz konusu olduğunda, davranış tekniği benzer sorunlara sahiptir (Birant, 2011).

3. Monetary

RFM analizindeki "M", "Monetary" kelimesinin baş harfidir ve "Parasal" anlamına gelir. Bu, bir müşterinin belirli bir zaman diliminde bir şirkete ne kadar para harcadığını ifade eder. Yani, bir müşterinin toplam harcamalarını temsil eder. Bu ölçüm, müşterinin şirket için ne kadar değerli olduğunu belirlemeye yardımcı olur. Genellikle, daha yüksek parasal değere sahip müşteriler daha değerli kabul edilir çünkü daha fazla gelir sağlarlar.

Bir müşterinin parasal değeri, ne kadar para harcadıklarıyla belirlenir. Doğal eğilim, en çok para harcayan tüketicileri bunu yapmaya devam etmeleri için daha çok teşvik etmeye yöneliktir. Bu, pazarlama ve müşteri hizmetleri yatırımlarında daha yüksek bir getiri sağlayabilirken, aynı zamanda her işlemde çok fazla harcama yapmayan sadık müşterileri kaybetme riskini de taşır (Birant, 2011).

IV. UYGULAMA

Bu çalışma, bir tekstil perakendecisinin gerçek müşteri verilerini temel alarak, firmanın müşteri satın alma davranışlarını daha derinlemesine ve doğru bir şekilde analiz edebilmesi amacıyla hazırlanmıştır. Bu analizler, firmanın pazarlama stratejilerini ve faaliyetlerini daha bilinçli ve etkili bir şekilde yönlendirebilmesi için kritik bir öneme sahiptir.

Bu bölüm, kullanılan veri setinin detaylı bir tanıtımını, veri seti üzerinde gerçekleştirilen ön işleme süreçlerini, kümeleme algoritmalarının nasıl uygulandığını, bu algoritmaların sonuçlarının nasıl değerlendirildiğini ve elde edilen müşteri segmentlerine dair yapılan analizleri ve yorumları içermektedir.

Çalışmanın ana eksenini oluşturan RFM analizi, K-ortalama ve bulanık C-ortalama algoritmaları, Python programlama dili ile geliştirilmiştir. Bu süreçte kullanılan Python kütüphanelerinin amacı aşağıda verilmiştir:

- **numpy**: Veri üzerinde matematiksel işlemler gerçekleştirmek için kullanılacaktır.
- **pandas**: Veri setlerini işlemek, temizlemek ve analiz etmek için kullanılacaktır. Tablo formatında veri manipülasyonu sağlar.
- **sklearn**: Kümeleme algoritmalarının uygulanması ve model değerlendirme metriklerinin elde edilmesi için kullanılacaktır.
- **seaborn ve matplotlib**: Veri görselleştirme işlemleri için kullanılacak ve elde edilen sonuçların grafiksel olarak sunulmasını sağlayacaktır.
- **skfuzzy**: Bulanık mantıkla çalışmak ve bulanık C-ortalama algoritmasını uygulamak için kullanılacaktır.
- **scipy**: İleri düzey matematiksel işlemler ve istatistiksel analizler için kullanılacaktır.
- **plotly**: Dinamik ve etkileşimli grafikler oluşturmak için kullanılacaktır.

Bu kütüphaneler, çalışmanın her aşamasında verinin doğru bir şekilde işlenmesi, analiz edilmesi ve sonuçların etkili bir şekilde sunulması için kritik bir rol oynamaktadır.

A. Veri Seti ve Veri Ön İşleme

Bu çalışma, Türkiye'de aktif olarak faaliyet gösteren ve sektörün önde gelen isimlerinden biri olan bir hazır giyim perakende firmasının gerçek verilerini kullanmaktadır. Firma, geniş bir coğrafi alanda, 200'e yakın mağaza ile hizmet vermektedir ve aynı zamanda güçlü bir internet mağazası ile de online alışveriş imkânı sunmaktadır.

Firmadan alınan bu veriler hem fiziksel mağazalarda hem de internet mağazasında gerçekleşen satışları kapsamaktadır. Bu veriler, hangi müşterinin hangi ürünü, kaç adet ve hangi fiyatla satın aldığı gibi temel satış bilgilerini içermektedir. Bu bilgiler, müşteri davranışlarını anlamak ve segmentasyon yapmak için oldukça değerlidir. Verilerin gizliliği ve müşteri bilgilerinin korunması açısından, tüm veriler maskelenmiştir ve sadece firma tarafından bilinen benzersiz kodlar ile paylaşılmıştır. Bu hem müşteri gizliliğini korurken hem de verilerin analiz edilmesi için gerekli olan bilgileri sağlar.

Bu çalışmanın amacı, bu verileri kullanarak müşteri segmentasyonunu oluşturmak, oluşturulan segmentlere ait müşterileri analiz etmek ve bu analizlerin sonuçlarını kampanya ve pazarlama faaliyetlerinde kullanmaktır. Bu, firmanın müşteri ilişkilerini daha etkin bir şekilde yönetmesine ve pazarlama stratejilerini daha hedef odaklı hale getirmesine yardımcı olacaktır.

Firma tarafından paylaşılan ham verilere ait örnek bir tablo aşağıda paylaşılmıştır. Bu tablo, analiz sürecinde kullanılan verilerin bir örneğini sunmaktadır ve bu verilerin nasıl kullanıldığını ve analiz edildiğini daha iyi anlamamıza yardımcı olacaktır.

Müşteri Kodu	Ürün Kodu	Miktarı	Fatura Numarası	Hesap Tarihi	Fatura Tarihi	Mağaza Kodu	Satış Fiyatı	ANA GRUP	ALT GRUP
M1	U1	1	F1	2015-06-10	2021-01-14 00:00:00	Store1	89,99	D0	DJ
M2	U2	1	F2	2015-06-04	2021-01-11 00:00:00	Store2	23,85	C1	C1
M3	U3	1	F3	2015-06-05	2021-01-12 00:00:00	Store3	223,99	C0	CB
M4	U4	1	F4	2008-06-30	2021-01-15 00:00:00	Store4	200,45	D0	DD
M5	U5	1	F5	2015-06-05	2021-01-12 00:00:00	Store5	223,99	C0	CB
M6	U6	1	F6	2010-07-22	2021-01-07 00:00:00	Store6	89,99	D0	DF
M7	U7	1	F7	2010-07-22	2021-01-08 00:00:00	Store7	59,99	N0	NC
M8	U8	1	F8	2010-07-23	2021-01-04 00:00:00	Store8	79,99	40	4A
M9	U9	1	F9	2007-02-20	2021-01-06 00:00:00	Store9	89,99	C0	CB

Şekil 2. Firmadan alınan ham müşteri satış verileri.

Kolon Adı	Açıklama
Müşteri Kodu	Her bir müşteri özelinde verilen tekil kod.
Ürün Kodu	Her bir ürün özelinde verilen tekil kod.
Miktar	Satılan ürün adeti.
Fatura Numarası	Siparişe ait fatura numarası.
Hesap Tarihi	Müşterinin hesap oluşturma tarihi.
Fatura Tarihi	Fatura tarihi.
Mağaza Kodu	Satın alım yapılan mağaza kodu.
Satış Fiyatı	Müşteri satış fiyatı.
ANA GRUP	Satılan ürünün bulunduğu hiyerarşiye ait ana grup kodu.
ALT GRUP	Satılan ürünün bulunduğu hiyerarşiye ait alt grup kodu.

Şekil 3. Müşteri ham verisine ait kolon isimleri veri açıklamaları.

Müşteri segmentasyonu süreci, verinin bir dizi ön işleme adımından geçirilmesiyle başlayacaktır. Bu ön işleme adımları, verinin analize uygun bir formata getirilmesini ve tutarsızlığın giderilmesini içerir. Bu aşama, verinin temizliğini ve kalitesini sağlamak için kritik öneme sahiptir, çünkü bu faktörler sonuçların doğruluğunu ve güvenilirliğini doğrudan etkileyecektir.

Ön işleme süreci kapsamında, müşteri analizi için oldukça önemli olan Sıklık, Frekans ve Harcama Tutarı değerleri hesaplanmıştır. Bu değerler, bir müşterinin bir işletmeyle olan etkileşimini ve değerini ölçmek için kullanılır.

- **Sıklık (Recency):** Bu değer, bir müşterinin en son ne zaman satın alma işlemi gerçekleştirdiğini gösterir. Örneğin, bir müşterinin son satın almasının üzerinden uzun bir süre geçiyse, bu müşterinin işletmeyle olan ilişkisinin zayıflamış olabileceği anlamına gelir.
- **Frekans (Frequency):** Frekans, belirli bir zaman dilimi içerisinde bir müşterinin ne kadar sıklıkla satın alma işlemi gerçekleştirdiğini gösterir. Bu çalışmada, Frekans değerini temsil etmek için veri setinde bulunan "OrderCount" yani sipariş verme miktarı değeri kullanılmıştır. Ancak, veri

setindeki müşterilerin satın alma sıklığı genel olarak düşük. Bu nedenle, frekans değerinin kümelenme analizindeki etkisi, modele beklenen katkıyı sağlamıyor.

- **Harcama (Monetary):** Harcama değeri, bir müşterinin toplamda ne kadar harcama yaptığını gösterir. Ancak, zamanla para biriminin değeri değişebilir. Bu nedenle, her bir müşterinin satın alımı için Harcama değerini daha tutarlı bir şekilde hesaplamak amacıyla, Türkiye İstatistik Kurumu (TÜİK) tarafından geçmişten günümüze aylık olarak açıklanan enflasyon oranlarına başvurulmuştur. Bu enflasyon oranları, satış fiyatlarına geçmişe dönük olarak uygulanarak, zamanla değişen para biriminin değerini dikkate alacak şekilde düzeltilmiştir. Bu yaklaşım, daha gerçekçi ve tutarlı bir Harcama değeri oluşturmayı hedeflemektedir.

Bu iyileştirmeler, müşteri analizinin daha doğru ve etkili bir şekilde yapılmasına yardımcı olacaktır.

Ön işlemeden geçirildikten sonra verinin deseni aşağıdaki gibi olmuştur.

Müşteri Kodu	recency	monetary	frequency	receipt_date_max	receipt_date_min	order_count	time_gap
M1	330	3248,482127	0,01754386	2022-04-16 00:00:00	2021-08-31 00:00:00	4	228
M2	1	57114,66313	0,042721519	2023-03-11 00:00:00	2021-06-17 00:00:00	27	632
M3	2	22969,85177	0,041450777	2023-03-10 00:00:00	2021-01-27 00:00:00	32	772
M4	20	22598,49523	0,023578363	2023-02-20 00:00:00	2021-03-01 00:00:00	17	721
M5	141	14213,04377	0,029644269	2022-10-22 00:00:00	2021-06-03 00:00:00	15	506
M6	546	1560,907832	0,020661157	2021-09-12 00:00:00	2021-01-13 00:00:00	5	242
M7	589	7185,102242	0,082125604	2021-07-31 00:00:00	2021-01-05 00:00:00	17	207
M8	113	502,8570031	0,03030303	2022-11-19 00:00:00	2022-09-14 00:00:00	2	66
M9	6	22378,97907	0,051175657	2023-03-06 00:00:00	2021-03-13 00:00:00	37	723
M10	443	4604,979074	0,020080321	2021-12-24 00:00:00	2021-04-19 00:00:00	5	249

Şekil 4. Ön işlemeden geçirildikten sonra müşteri veri deseni.

Ön işleme süreci sonrasında yukarıda deseni verilen veri için K-ortalama ve bulanık C-ortalama kümeleme algoritmaları kullanılarak müşteriler farklı kümelere ayrılacaktır. Kümelere ayırma işlemi sonrasında ise kümelerin başarımları değerlendirilecektir. Yapılacak değerlendirme de Silhouette indeks kullanılacaktır. Bu algoritmalara ve yöntemlere ilişkin bilgiler alt başlıklar halinde verilmiştir.

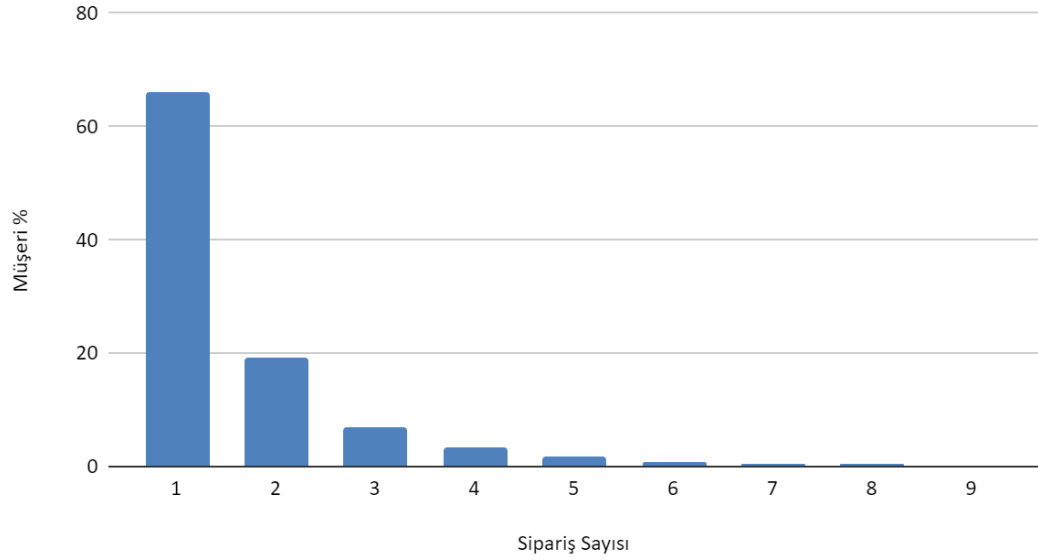
B. Kümleme Algoritmalarının Uygulanması

Bu çalışma kapsamında, veri ön işleme sürecinden geçirilmiş ve müşterilere ait RFM değerleri oluşturularak kümeleme için hazırlanmış bir veri seti üzerinde durulacaktır. Bu bölümde, ön işleme sürecinden geçirilen verilerin K-ortalama ve bulanık C-ortalama algoritmaları kullanılarak nasıl kümelere ayrıldığı ve bu kümelerin Silhouette Index metriği ile nasıl değerlendirildiği detaylı bir şekilde anlatılacaktır. Bu değerlendirme süreci, en başarılı kümeleme sonucunu belirlememize yardımcı olacaktır.

Firmadan alınan veri seti, 3 Ocak 2021 ve 12 Mart 2023 tarihleri arasındaki satış ve müşteri bilgilerini içermektedir. Bu veri seti incelendiğinde, toplamda 1.907.413 tekil müşteri olduğu görülmektedir. Ancak, daha etkili ve anlamlı kümeler oluşturabilmek için, bu iki yıllık dönem içerisinde sadece bir veya iki kez alışveriş yapan müşteriler veri setinden çıkarılmıştır. Bu dönem içerisinde sadece bir kez alışveriş yapmış olan tekil müşteri sayısı 1.260.324, sadece iki kez alışveriş yapmış olan müşteri sayısı ise 363.451'dir. Bu müşterilerin çıkarılmasının ardından, RFM verileri oluşturulan ve kümeleme analizine dahil edilen müşteri sayısı 283.638 tekil kişiye düşmüştür. Bu çalışmada gerçekleştirilen tüm analizler, bu 283.638 tekil müşteri üzerinden yapılmıştır. Bu süreç, veri setinin daha yönetilebilir bir boyuta indirgenmesini ve kümeleme analizinin daha anlamlı ve yorumlanabilir sonuçlar üretmesini sağlamıştır.

Aşağıda müşterilerin sipariş sayısı dağılımına göre grafiksel bir gösterim hazırlanmıştır.

Müşeri %-Sipariş Sayısı



Şekil 5. Sipariş sayısının müşteriye oranı

283.638 tekil müşteri verisiyle çalıştırılan k-ortalama ve bulanık c-ortalama algoritmalarının kümeleme başarımları Silhouette indeks ile değerlendirilmiştir. Her iki algoritmada farklı sayıda kümeler için tekrar çalıştırılmış ve küme sayılarındaki değişimin indeks değeri değerlendirilmiştir.

Her iki algoritmanın da farklı küme sayıları ile alınan en iyi sonuçlar aşağıdaki tabloda verilmiştir.

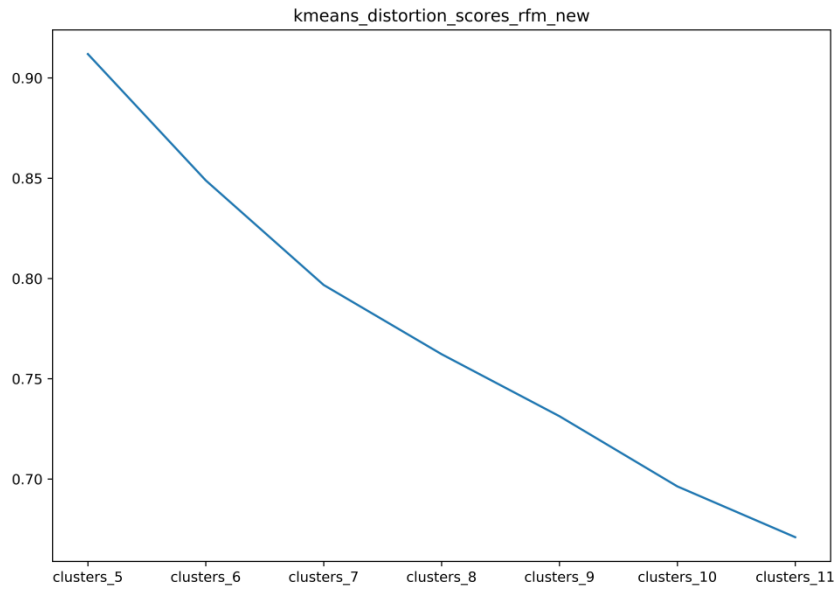
	5	6	7	8	9	10	11
kmeans	0,29543 7	0,30282 7	0,30644 9	0,30928 5	0,30520 3	0,31000 2	0,31052 5
fcmeans	0,28756 3	0,29004 2	0,29992 8	0,29975 1	0,29853 5	0,29410 6	0,29999 9

Şekil 6. K-ortalama ve bulanık c-ortalama algoritmalarının küme başarımlarının karşılaştırılması

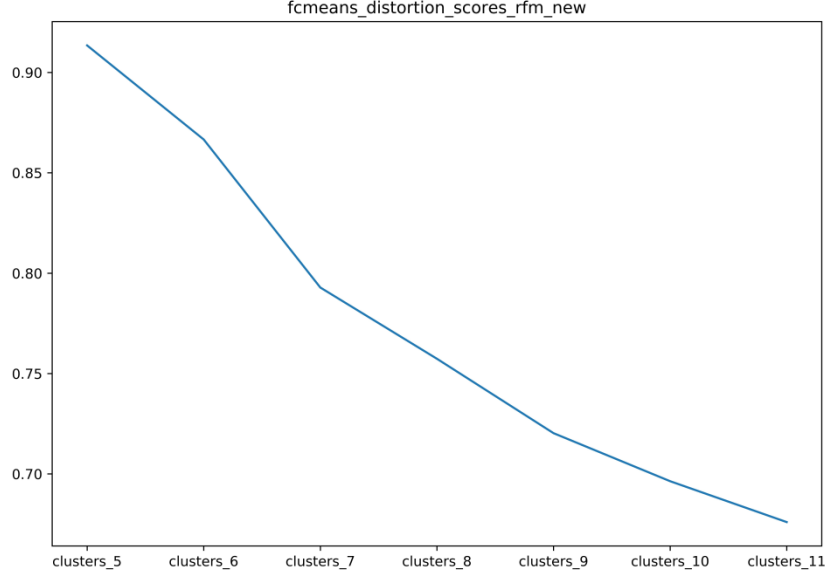
Tabloda belirtilen değerlerin incelenmesi sonucunda, K-ortalama algoritması için en uygun küme sayısının 8 olduğu sonucuna varılmıştır. Bu sonuç, başarımların 8 kümeli çözüm sonrasında düşüş göstermesi temel alınarak belirlenmiştir. Başka bir deyişle, 8 kümeli çözüm, algoritmanın başarımlarını maksimize ederken, daha fazla küme sayısı algoritmanın başarımlarını düşürmüştür. Bu nedenle, K-ortalama algoritması için en uygun küme sayısı 8 olarak belirlenmiştir. Öte yandan, bulanık C-ortalama algoritması için en uygun küme sayısı 5 olarak belirlenmiştir. Bu karar, başarımların 5 kümeli

özüm sonrasında düşüş göstermesi temel alınarak verilmiştir. Yani, 5 kümeli çözüml, bulanık C-ortalama algoritmasının başarımını maksimize ederken, daha fazla küme sayısı algoritmanın başarımını düşürmüştür. Bu nedenle, bulanık C-ortalama algoritması için en uygun küme sayısı 5 olarak belirlenmiştir. Bu seçimler, her iki algoritmanın başarımını maksimize etmek için yapılmıştır. Başarım değerlerinin düşüş gösterdiği küme sayısı, her iki algoritma için en uygun küme sayısını belirlememize yardımcı olmuştur. Bu yaklaşım, hem K-ortalama hem de bulanık C-ortalama algoritmalarının en etkili şekilde kullanılmasını sağlamıştır.

Küme sayısının seçimi noktasında küme içi uzaklık değeri de değerlendirilmiştir. Hem K-ortalama hem de bulanık C-ortalama için küme içi uzaklık değerleri aşağıda verilmiştir.



Şekil 7. K-ortalama algoritmasının farklı küme sayılarının toplam küme içi uzaklıkları



Şekil 8. Bulanık C-ortalama algoritmasının farklı küme sayılarının toplam küme içi uzaklıkları

Başarım değerinin optimum olarak belirlendiği K-ortalama 8 ve bulanık C-ortalama 5 kümeleri üzerinden yapılmasına bu performans kriterlerine bakılarak karar verilmiştir.

1. K-ortalama ile RFM Analizi

Kümeleme analizinin başarımını değerlendirmek için kullanılan Silhouette indeksi ve kümelerin genel başarım değerleri dikkate alındığında, K-ortalama algoritması için en uygun küme sayısının sekiz olduğu sonucuna varılmıştır. Bu sonuç, hem Silhouette indeksinin en yüksek değerini veren küme sayısı hem de genel başarım değerlerinin en yüksek olduğu küme sayısı olarak belirlenmiştir. Bu sonuca dayanarak, sekizli küme çözümü üzerinde daha detaylı analiz ve değerlendirme çalışmaları gerçekleştirilmiştir. Bu analizler, sekizli küme çözümünün her bir kümesinin karakteristik özelliklerini, müşteri davranışlarını ve alışveriş alışkanlıklarını daha iyi anlamamızı sağlamıştır. Sekizli küme çözümüne ait veriler ve bu verilerin analiz sonuçları aşağıdaki tabloda sunulmuştur. Bu tablo, her bir kümenin RFM değerlerini, müşteri sayısını ve diğer önemli özelliklerini içermektedir.

No	monetary	monetary_ran	order_coun	order_count_ran	recenc	recency_ran
		k	t	k	y	k
0	6.258,32	5	6,061	7	298,54	5
1	2.152,45	2	3,470	2	439,11	8
2	8.876,04	7	3,673	5	361,86	6
3	2.952,34	3	3,480	3	405,02	7
4	2.103,06	1	3,516	4	273,81	3
5	4.008,94	4	3,771	6	70,516	1
6	7.160,83	6	3,326	1	289,08	4
7	18.410,77	8	9,522	8	98,572	2

Şekil 9. K-ortalama algoritması sekizli küme sonuçları

K-ortalama algoritması ile yapılan RFM kümelemesinde ortaya çıkan müşteri kümeleri özelinde müşterilerin davranışlarına ve firma ile olan alışveriş süreçlerine yönelik yorumlar yapılabilmektedir. Bu kümelerin en dikkat çekenlerinden biri de 7 numaralı küme olmuştur. Bu küme 18.410 birimlik harcama ortalamasına sahip ve sıklığı 98 olarak görülen sadık ve yüksek harcama yapan müşteri grubunu ifade etmektedir. Bu grup aynı zamanda en yüksek sipariş sıklığı ortalaması olan 9,5 oranına sahiptir.

1 numaralı küme, dikkatlice incelendiğinde, sıklık değeri 439 olan bir müşteri grubunu temsil etmektedir. Bu değer, bu grubun potansiyel olarak kaybedilmiş müşterileri içerdiğini göstermektedir. Ancak, bu grubun harcama ortalamasına baktığımızda, oldukça düşük bir değer olan 2.152 birimi görüyoruz. Aynı zamanda, bu gruptaki müşterilerin sipariş ortalaması da 3.470 olarak belirlenmiştir. Bu, bu grubun düşük harcama yapan ve nadiren sipariş veren müşterileri içerdiğini göstermektedir. Bu küme özelinde müşterilerin verdikleri siparişler ürün grubu bazında incelenebilir ve müşterilerin hangi ürün gruplarına eğitimleri oldukları değerlendirilebilir. Eğer müşteriler belirli ucuz ürün gruplarına eğitim gösteriyor bu yüzden daha düşük bir katkı sağlayan grupta bulunuyorlarsa bu ürünler özelinde pazarlama faaliyetleri yürütülmesi faydalı olabilir.

Ancak, 1 numaralı kümeden daha kritik bir grup olan 2 numaralı küme de bulunmaktadır. 2 numaralı küme, sıklık değeri 361 olan bir müşteri grubunu temsil etmektedir. Bu grup, son alışverişlerinin üzerinden oldukça uzun bir süre geçmiş olan ve bu nedenle potansiyel olarak kaybedilmiş müşterileri içerir. Ancak, bu grubun harcama ortalamasına baktığımızda, 8.876 birimlik bir değer görüyoruz. Bu, bu grubun yüksek harcama yapan müşterileri içerdiğini göstermektedir. Bu iki grup arasındaki bu önemli fark, pazarlama stratejilerinin belirlenmesinde önemli bir rol oynar. 2 numaralı grubun yüksek harcama ortalaması, bu grubun tekrar alışveriş yapmalarını sağlamak için özel pazarlama stratejileri geliştirilmesi gerektiğini gösterir. Eğer bu müşterilerin tekrar alışveriş yapmaları sağlanabilirse, firmanın potansiyel kazancı 1 numaralı kümeyle göre çok daha yüksek olacaktır. Bu, müşteri ilişkileri yönetimi ve müşteri geri kazanma stratejilerinin önemini vurgular.

K-ortalama algoritması tarafından oluşturulan kümeler arasında, özellikle 5 numaralı grup dikkat çekmektedir. Bu müşteri grubu, firma bünyesinde en yakın zamanda alışveriş yapmış olan müşterileri temsil etmektedir. Bu grubun son alışverişlerinin üzerinden geçen ortalama süre sadece 70 gün olarak belirlenmiştir. Bu, bu grubun müşterilerinin firma ile son derece güncel bir ilişkisi olduğunu göstermektedir. Aynı zamanda, bu müşteri grubunun sipariş ortalaması da 3,771 olarak belirlenmiştir. Bu, bu müşterilerin analiz dönemi boyunca en az ortalama 3,771 kez alışveriş yaptığını göstermektedir. Ancak, bu grubun harcama ortalamasının 4.008 birim olduğu görülmektedir. Bu, bu müşteri grubunun sadık ve yakın zamanda mağazalara gelmiş olmasına rağmen, harcama tutarlarının ortalamasının altında olduğunu göstermektedir. Bu analiz sonuçları, bu müşteri grubuna özel pazarlama stratejileri geliştirilmesi gerektiğini göstermektedir. Bu müşteriler, sadık ve yakın zamanda mağazalara gelmiş olmalarına rağmen, harcama tutarlarının ortalamasının altında olması nedeniyle, harcamalarını artırmaya yönelik özel pazarlama faaliyetleri ile hedeflenmelidir. Bu, firmanın bu müşteri grubunun harcama potansiyelini daha etkin bir şekilde değerlendirmesine ve bu müşterilerden daha yüksek bir gelir elde etmesine yardımcı olabilir.

K-ortalama algoritması kullanılarak oluşturulan 6 numaralı kümeyle detaylı bir inceleme yaptığımızda, bu segmentin firmanın değerli müşteri grubunu temsil

ettiğini gözlemleyebiliriz. Bu segmentteki toplam harcama 7.160 birim olarak belirlenmiştir. Bu, ilk bakışta ortalama bir harcama gibi görünse de bu harcamanın gerçekleştiği toplam sipariş sayısının 3,326 olduğunu göz önünde bulundurduğumuzda durum değişiklik gösteriyor. Bu, bu müşteri grubunun mağazayı ortalama 3,326 kez ziyaret ederek toplamda 7.160 birim harcama yaptığı anlamına gelir. Bu segmenti, benzer bir harcama değerine sahip olan 0 numaralı kümeye kıyasladığımızda, en belirgin farkın toplam harcamayı oluşturan sipariş sayısında olduğunu belirleyebiliriz. Ayrıca, bu müşteri grubunun son alışverişlerinden bu yana 289 gün geçtiğini gözlemliyoruz. Bu, sepet derinliği yüksek ve sipariş başına harcama miktarı oldukça fazla olan bu müşteri grubunun, veri ortalamasına göre uzun bir süre boyunca alışveriş yapmadığı anlamına gelir. Bu durum, bu potansiyel açısından zengin müşteri grubu için özel bir pazarlama stratejisi uygulanmazsa, ilerleyen dönemlerde bu müşterilerin firmadan uzaklaşabileceği riskini ortaya koymaktadır.

Alternatif bir yaklaşım olarak, 0 numaralı müşteri grubunu ele alabiliriz. Bu grup, 6.258 birimlik ortalama bir harcama tutarına sahip olup, analiz dönemi boyunca ortalama 6,061 sipariş vermiştir. Bu, bu grubun yüksek bir sipariş sıklığına sahip olduğunu göstermektedir. Ancak, bu müşterilerin son alışverişlerinin üzerinden ortalama 298 gün geçmiştir. Bu durum, bu müşteri grubunun alışveriş sıklığının yüksek olmasına rağmen, son zamanlarda firma ile etkileşime geçmediğini göstermektedir. Bu müşteri grubu da özellikle incelenmesi gereken bir grup olarak görülmektedir. Müşterilerin hem harcama tutarları ortalamaya yakın olduğu gibi hem de sipariş verme oranları yüksektir. Yani bu müşteri grubu ortalama 6 kez mağazalar veya internet mağazası üzerinden sipariş vermiş olmasına rağmen son 298 gündür herhangi bir alışveriş yapmamıştır.

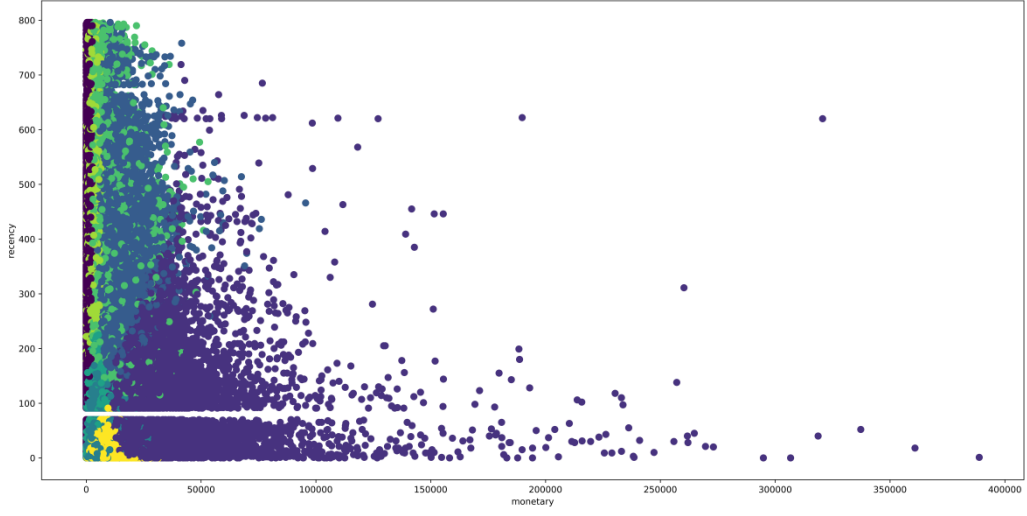
3 ve 4 numaralı kümeler arasındaki benzerlikleri ve farkları derinlemesine incelemek, stratejik pazarlama kararları için oldukça aydınlatıcı olabilir. İlk olarak, bu iki kümeye genel bir bakış attığımızda, her iki grubun da ortalama harcama tutarlarının genel veri setinin ortalamasının altında olduğunu görebiliriz. Bu, her iki grubun da daha tutucu veya seçici alışveriş yaptığını gösteriyor olabilir. 3 numaralı kümenin ortalama harcama toplamı 2.952 birimken, 4 numaralı küme için bu değer 2.103 birimdir. Bu, 4 numaralı kümeyi, veri

setindeki en düşük harcamaya sahip grup olarak öne çıkartmaktadır. Ancak, her iki grubun sipariş ortalamaları oldukça benzerdir: 3,48 ve 3,51. Bu, her iki grubun da benzer alışveriş sıklığına sahip olduğunu, ancak harcadıkları miktarın farklı olduğunu gösteriyor. Bu iki grubun harcama alışkanlıkları genel olarak birbirine yakın olsa da mağazalardan yaptıkları son alışverişten bu yana geçen süre, bu iki grubu birbirinden ayıran en belirgin özelliktir. 3 numaralı gruptaki müşterilerin son alışverişlerinden bu yana 405 gün geçmiştir. Bu süre, genel veri setinin ortalamasının üzerinde olduğundan, bu müşteri grubunun potansiyel olarak kaybedilmiş müşterileri temsil ettiği söylenebilir. Öte yandan, 4 numaralı grupta bu süre 273 gün olarak belirlenmiştir. Bu, bu grubun daha yakın bir zamanda alışveriş yaptığını, ancak yine de potansiyel bir risk taşıdığını göstermektedir. Bu analizler ışığında, 3 numaralı grup için firmanın acil bir şekilde müşteri geri kazanım stratejileri geliştirmesi ve bu müşterilere özel teklifler sunması önerilir. 4 numaralı grup içinse, bu müşterilerin ilgisini sürdürebilmek ve onları daha sık alışveriş yapmaya teşvik edebilmek için özelleştirilmiş pazarlama kampanyaları ve teklifler oluşturulması faydalı olacaktır. Eğer bu adımlar atılmazsa, her iki grup da zamanla firmanın müşteri tabanından kaybolabilir.

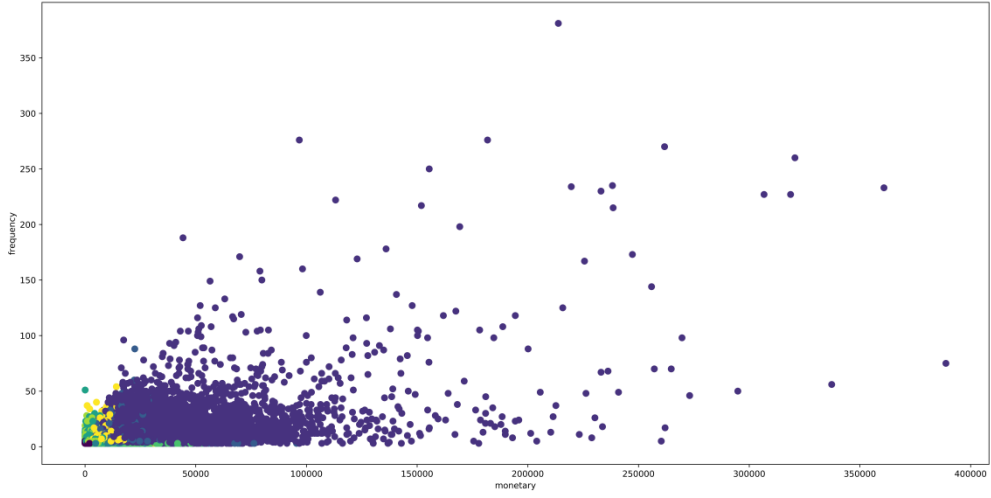
Kümelerin oluşturulmasında RFM değerleri temel alınarak bir segmentasyon stratejisi benimsenmiştir. RFM değerleri, müşterilerin alışveriş sıklığını, en son ne zaman alışveriş yaptıklarını ve genel olarak ne kadar harcama yaptıklarını temsil eder. Bu üç değer, müşteri davranışlarını ve alışveriş alışkanlıklarını daha iyi anlamak için kritik öneme sahiptir. Bulanık C-ortalama yönteminde olduğu gibi, K-ortalama algoritması da bu RFM değerlerini kullanarak müşterileri belirli segmentlere ayırmıştır. Bu segmentasyon süreci, her bir müşterinin hangi kümeye daha yakın olduğunu belirlemek için RFM değerlerinin birbirleriyle olan ilişkilerini dikkate almıştır.

Segmentasyon sonucunda oluşturulan kümelerin dağılımlarını daha iyi anlamak ve görsel olarak temsil etmek için çeşitli görselleştirme teknikleri kullanılmıştır. Bu görselleştirmeler, kümelerin birbirlerine göre konumlandırılmasını, hangi kümelerin benzer veya farklı özelliklere sahip olduğunu ve genel olarak segmentasyonun ne kadar başarılı olduğunu değerlendirmek için oldukça faydalıdır. Bu görselleştirmeler, metnin ilerleyen

bölümlerinde detaylı bir şekilde sunulmuştur ve bu görseller, kümelerin dağılımını anlama ve yorumlama sürecinde kritik bir rol oynamaktadır.



Şekil 10. Sıklık ve harcama tutarı değerlerinin küme dağılımı



Şekil 11. Frekans ve harcama tutarı değerlerinin küme dağılımı

2. Bulanık C-Ortalama ile RFM Analizi

Başarımların değerlendirmesi sonrasında karar verilen 5 kümeli bulanık c-ortalama algoritması ile üretilen kümelere ait detay aşağıdaki tabloda verilmiştir.

No	monetary	monetary_rank	order_count	order_count_rank	recency	recency_rank
0	2.210,24	1	3,671	3	317,527	3
1	5.417,75	4	3,545	1	399,086	5
2	13.049,09	5	7,587	5	130,964	1
3	5.395,86	3	3,864	4	181,43	2
4	5.172,79	2	3,562	2	321,103	4

Şekil 12. Bulanık C-ortalama küme sonuçları

Bu beşli küme çözümüne özel olarak müşteri dağılımını incelediğimizde, 2 numaralı kümenin en değerli müşteri grubunu oluşturduğunu gözlemleyebiliriz. 2 numaralı küme, müşteri harcamaları ve sipariş sayısı bakımından diğer kümelere belirgin bir şekilde ayrılıyor. 2 numaralı küme içerisinde yer alan müşterilerin, incelenen veri süresince ortalama 13.049 birimlik bir harcama yaptığını görüyoruz. Ayrıca, bu müşteri grubunun ortalama olarak bu süre zarfında 7,5 sipariş verdiği belirlenmiştir. Bu, bu kümenin hem yüksek harcama yapan hem de sık sipariş veren müşterilerden oluştuğunu göstermektedir. Bununla birlikte, aynı müşteri grubunun sıklık değerine baktığımızda, en düşük değer olan 130'u görüyoruz. Bu, 2 numaralı kümenin, firmanın en sadık müşteri grubunu oluşturduğunu göstermektedir. Yani, bu müşteri grubu, firmanın en değerli müşterilerini oluşturuyor. Hem yüksek miktarda harcama yapmaları hem de sık sipariş verme eğilimleri, bu grubun firmanın gelirlerine önemli bir katkı sağladığını göstermektedir. Bu bulgular, firmanın pazarlama ve satış stratejilerini şekillendirmede önemli bir rol oynayabilir. Özellikle, 2 numaralı küme gibi yüksek değerli müşteri gruplarının ihtiyaçlarına ve tercihlerine odaklanmak, firmanın gelirlerini artırmada etkili olabilir.

Beşli küme çözümü üzerinde dikkatle durduğumuzda, 1 numaralı kümenin de özellikle odaklanmamız gereken bir diğer önemli grup olduğunu belirtmek gerekir. 1 numaralı küme, harcama tutarı bakımından ortalama bir değere sahip olan bir müşteri grubunu temsil eder. Bu kümenin müşterileri, incelenen veri süresince ortalama 5.417 birimlik bir harcama yapmışlardır. Bu miktar, bu grubun firmanın gelirlerine ortalama bir katkı sağladığını göstermektedir. Ancak, bu

kümenin sıklık değerine baktığımızda, ortalama 399 değerini görüyoruz. Bu değer, bu müşteri grubunun son alışverişlerinin üzerinden oldukça uzun bir süre geçtiğini göstermektedir. Bu durum, bu müşterilerin potansiyel olarak kayıp müşteri olarak kabul edilebileceğini göstermektedir. Yani, bu müşteriler bir süre önce firmanın ürünlerini satın almayı bırakmış olabilirler. Bu bulgular, firmanın pazarlama stratejilerini şekillendirmede önemli bir rol oynayabilir. Özellikle, 1 numaralı küme gibi potansiyel kayıp müşteri gruplarına odaklanmak, firmanın gelirlerini artırmada etkili olabilir. Bu müşteri grubuna yönelik özel pazarlama faaliyetleri yürütmek ve bu müşterileri tekrar firmanın ürünlerini satın almaya teşvik etmek, firmanın gelirlerini artırmada önemli bir rol oynayabilir.

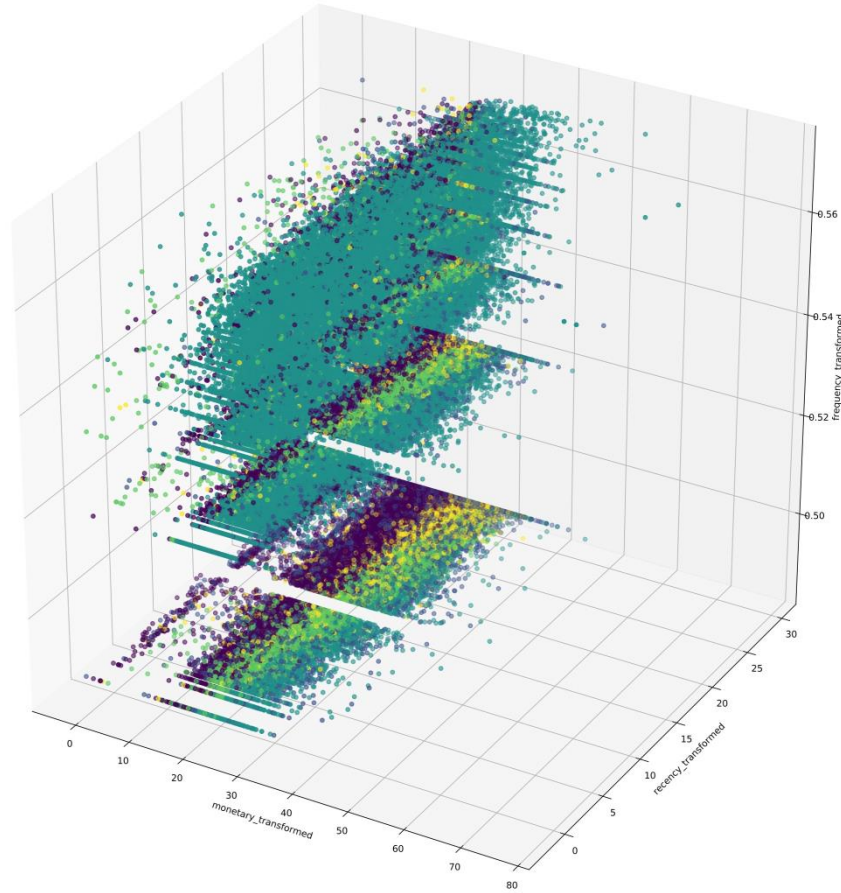
3 numaralı kümeye odaklandığımızda, bu grubun ortalama 5.395 birim harcadığını gözlemliyoruz. Bu değer, en yüksek harcamaya sahip müşteri grubunu takip eden oldukça iyi bir rakamdır ve firmanın gelirinde önemli bir paya sahip olabileceğini gösterir. Aynı zamanda, bu müşteri grubunun son alışverişlerinden bu yana geçen süre ortalama 181 gün olarak belirlenmiştir. Bu süre, onları ikinci en sadık müşteri grubu yapmaktadır. Sadakatleri, firmanın bu segment üzerinde özel pazarlama faaliyetleri yürütmesi için bir fırsat olabilir. Sipariş ortalamaları 3,86 gibi yüksek bir değerle karşılaşıyoruz, bu da onların sık sık alışveriş yaptığını gösteriyor. Bu segmentin geçmiş alışveriş trendlerini dikkate alarak, harcamalarını artırmak için özel pazarlama stratejileri geliştirmek, firmanın gelirini daha da artırabilir.

4 numaralı müşteri grubunu incelediğimizde, harcama tutarı ve sipariş ortalaması açısından 3 numaralı gruba oldukça benzer olduklarını görüyoruz. Ancak, bu iki grubu birbirinden ayıran kritik bir fark bulunmaktadır: recency değeri. 4 numaralı grubun son alışverişlerinden bu yana geçen süre tam 321 gündür. Bu uzun süre, firmanın bu müşteri grubuyla etkileşimde bulunmadığını veya etkileşimin yetersiz olduğunu gösteriyor. Ancak, bu grubun firmanın değerli bir müşteri segmenti olduğunu işaret eden 5.172 birimlik iyi bir harcama ortalamasına ve 3,52 sipariş ortalamasına sahip olması, onları tekrar aktif hale getirebilmek için özel stratejiler geliştirmenin önemini vurgular.

Segmentasyonumuzun son kümesi olan 0 numaralı küme, diğer kümelerden farklı bir profil çiziyor. Harcama tutarı açısından en düşük harcamalı müşteri grubunu oluşturuyor. Bu grubun harcama tutarı 2.210 birimdir, bu da veri

setindeki en düşük harcamaya işaret eder. Ancak, bu müşteri grubunun potansiyelini göz ardı etmek büyük bir hata olurdu. Sipariş ortalamaları 3.67 ile diğer segmentlere benzer olsa da bu değer ve harcama tutarının kombinasyonu, bu müşterilerin ziyaret başına düşük sepet ortalamalarına sahip olduğunu gösterir. Bu, firmanın bu müşteri grubuna yönelik özel kampanyalar ve yüksek indirimli ürünlerle müşteriyi tekrar firmanın bünyesine çekme potansiyeline sahip olduğunu gösterir.

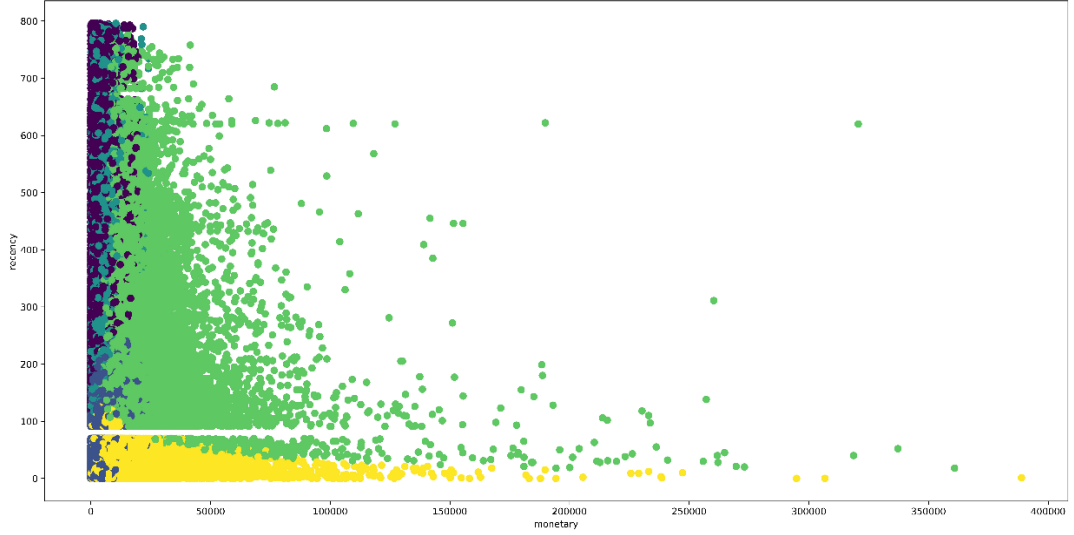
Aşağıda bu beşli küme grubu özelinde üç boyutlu bir gösterim ile kümelerin dağılımı bulunmaktadır.



Şekil 13. Bulanık C-ortalama kümelerinin üç boyutlu şekillendirmesi

Kümeleme analizinin bir parçası olarak, kümeler içerisinde hesaplanan RFM değerlerinin birbirleriyle olan ilişkileri de detaylı bir şekilde incelenmiştir. Bu değerlerin birbirleriyle olan ilişkileri, verinin yorumlanması ve müşteri dağılımının daha derinlemesine anlaşılması açısından büyük önem taşımaktadır. RFM değerleri, müşteri davranışlarının ve alışveriş alışkanlıklarının daha iyi anlaşılmasına yardımcı olur ve bu da pazarlama stratejilerinin daha etkili bir

şekilde belirlenmesine olanak sağlar. Bu beşli küme çözümüne özel olarak, müşteriler için hesaplanan RFM değerlerinin birbirleriyle olan ilişkileri de görselleştirilmiştir. Bu görselleştirme, her bir kümenin karakteristik özelliklerini ve bu özelliklerin nasıl bir araya geldiğini daha iyi anlamamızı sağlar. Bu kapsamda, aşağıda RFM değerlerinin birbirleriyle olan ilişkilerini gösteren görsellere yer verilmiştir.

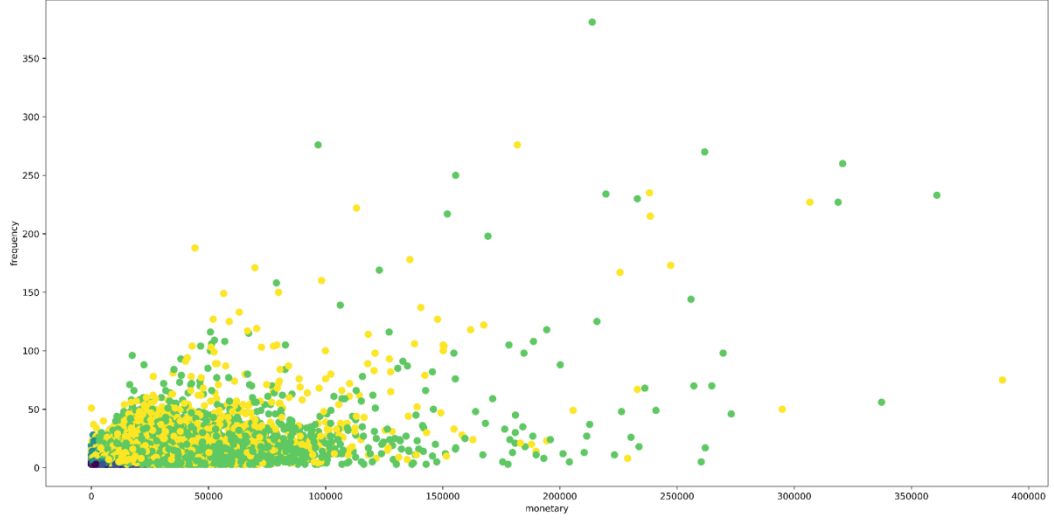


Şekil 14. Sıklık ve harcama miktarı değerlerinin küme dağılımı

Yukarıdaki görselde görüldüğü üzere, müşterilerin büyük bir çoğunluğu 0 ile 50.000 birim arasında harcama yapmıştır. Ancak, bu değer üzerinde harcama yapan ve alışveriş sıklığı yüksek olan müşterilerin de olduğunu gözlemlemekteyiz.

Küme dağılımlarının incelenmesinin faydalı olduğu bir başka başlık da frekans ve harcama tutarı verilerinin küme dağılımı olacaktır.

Bu dağılım da aşağıdaki görselde verilmiştir.



Şekil 15. Frekans ve harcama tutarı değerlerinin küme dağılımı

V.SONUÇ

Bu çalışmada, 1.907.413 adet tekil müşteriye ait 2 yıllık döneme ilişkin gerçek veriler üzerinden detaylı bir RFM analizi gerçekleştirilmiştir. Bu analiz sürecinde, müşteri verileri K-ortalama ve bulanık C-ortalama algoritmaları kullanılarak belirli kümeler halinde segmentlere ayrılmıştır. Bu iki algoritmanın kümeleme performansları dikkatlice değerlendirilmiş ve analiz edilmiştir. Her iki algoritma için de en optimum küme sayıları belirlenmiştir. Bu optimum küme sayılarına dayanarak, belirlenen sonuçlar üzerinde detaylı analizler yapılmıştır. Bu analizler, her iki algoritmanın da firmanın müşteri tabanını başarılı bir şekilde segmentlere ayırdığını ve elde edilen verinin, müşteri satış davranışlarını analiz etme ve pazarlama faaliyetlerini yönlendirme açısından son derece değerli olduğunu ortaya koymuştur. Bununla birlikte, bulanık C-ortalama algoritması ile belirlenen beşli kümeleme grubu ile K-ortalama algoritması tarafından üretilen sekizli küme karşılaştırıldığında, sekizli kümenin müşterilere yönelik daha derinlemesine bir analiz sunduğu görülmüştür. Her iki algoritma tarafından üretilen kümeler detaylı bir şekilde incelendiğinde, firmanın hem kaybettiği varsayılan hem de kaybetmeye yakın olan müşterileri ile sadık müşterilerini analiz etme noktasında önemli sonuçlar elde edildiği görülmektedir. Bu kümeler üzerinden müşterilere odaklanan ve müşteri deneyimini iyileştirmeye yönelik pazarlama stratejileri geliştirilmesi, firmanın kısa sürede önemli geri dönüşler elde etme potansiyelini artırabilir. Bu durum, müşteri odaklı pazarlama ve müşteri deneyimi süreçlerinin firmalar için ne kadar kritik olduğunu bir kez daha göstermektedir.

Gelecekteki çalışmalarda, aykırı değer tespiti ve çıkarma konusunda daha gelişmiş yaklaşımların benimsenmesi büyük önem taşımaktadır. Özellikle Isolation Forests ve DBSCAN gibi modern teknikler, aykırı değerlerin tespit edilmesinde oldukça etkili olabilir. Bu teknikler, veri setindeki anomalileri daha hassas bir şekilde belirleyerek, analiz sonuçlarının doğruluğunu artırabilir.

Perakende sektörü, özellikle moda endüstrisi, diğer perakende pazarlarına kıyasla farklı bir satın alma davranışına sahiptir. Moda sektöründe, tüketicilerin satın alma davranışı genellikle daha dikkatli ve seçicidir. Bu, veri noktalarının genellikle minimum değere yakın olmasına neden olabilir. Bu özgül davranış, veri ön işleme ve dönüştürme süreçlerinin daha detaylı ve özenli bir şekilde ele alınmasını gerektirir. Kümeleme analizinde, farklı algoritmaların veya mevcut algoritmaların varyantlarının kullanılması da değerlendirilmelidir. Örneğin, K-medoidler gibi alternatif kümeleme yöntemleri, bu çalışmada kullanılan yöntemlere kıyasla farklı avantajlar sunabilir.

Son olarak, kümeleme performansının değerlendirilmesi için daha kapsamlı kriterlerin eklenmesi önerilmektedir. Özellikle Dunn veya Tao İndeksleri gibi iç kriterler, kümeleme sonuçlarının kalitesini daha objektif bir şekilde değerlendirmek için kullanılabilir. Bu tür indeksler, kümeler arasındaki benzerlik veya farklılık gibi ölçütleri dikkate alarak, kümeleme sonuçlarının ne kadar başarılı olduğunu belirlemeye yardımcı olabilir. Bu önerilerin benimsenmesi, gelecekteki çalışmaların daha kapsamlı, doğru ve etkili sonuçlar elde etmesine katkıda bulunabilir.

VI. KAYNAKÇA

KİTAPLAR

BURKOV, A. (2019). **The Hundred-Page Machine Learning Book**. Andriy Burkov

GOODFELLOW, I., BENGIO, Y., & COURVILLE, A. (2016). **Deep Learning**. Amsterdam University Press.

TRAPPENBERG, T. (2019). **Fundamentals of Machine Learning**. Oxford University Press.

DERGİLER

ABİRAMİ, M., & PATTABİRAMAN, V. (2016). Data Mining Approach for Intelligent Customer Behavior Analysis for a Retail Store. Proceedings of the 3rd **International Symposium on Big Data and Cloud Computing Challenges** (ISBCC – 16'), 283–291. https://doi.org/10.1007/978-3-319-30348-2_23

BARUS, O. P., NATHASYA, C., & PANGARİBUAN, J. J. (2023). The Implementation of RFM Analysis to Customer Profiling Using K-Means Clustering. **Mathematical Modelling of Engineering Problems**, 10(1).

BAYKASOĞLU, A., GÖLCÜK, İ., & ÖZSOYDAN, F. (2018). Improving fuzzy c-means clustering via quantum-enhanced weighted superposition attraction algorithm. **Hacettepe Journal of Mathematics and Statistics**, 48(3), 859-882. <https://doi.org/10.15672/hjms.2019.657>

BEZDEK, J. C., EHRLICH, R., & FULL, W. (1984). FCM: The fuzzy c-means clustering algorithm. **Computers & geosciences**, 10(2-3), 191-203.

- CHAN (2008) “Intelligent value-based customer segmentation method for campaign management: A case study of automobile retailer” **Expert Systems with Applications** Volume 34, Issue 4, May 2008, Pages 2754-2762 <https://doi.org/10.1016/j.eswa.2007.05.043>
- CHEN, Y. L., KUO, M. H., WU, S. Y., & TANG, K. (2009). Discovering recency, frequency, and monetary (RFM) sequential patterns from customers’ purchasing data. **Electronic Commerce Research and Applications**, 8(5), 241– 251. <https://doi.org/10.1016/j.elerap.2009.03.002>
- CHENG, C. H., & CHEN, Y. S. (2009). Classifying the segmentation of customer value via RFM model and RS theory. **Expert Systems with Applications**, 36(3), 4176–4184. <https://doi.org/10.1016/j.eswa.2008.04.003>
- CHRİSTY, A. J., UMAMAKESWARİ, A., PRIYATHARSİNİ, L., & NEYAA, A. (2018). RFM ranking—An effective approach to customer segmentation. **Journal of King Saud University-Computer and Information Sciences**
- CHRİSTY, A.J. (2018). RFM ranking – An effective approach to customer segmentation. **Journal of King Saud University – Computer and Information Sciences**. <https://doi.org/10.1016/j.jksuci.2018.09.004>
- COUSSEMENT, K., VAN DEN BOSSCHE, F. A., & DE BOCK, K. W. (2014). Data accuracy’s impact on segmentation performance: Benchmarking RFM analysis, logistic regression, and decision trees. **Journal of Business Research**, 67(1), 2751– 2758. <https://doi.org/10.1016/j.jbusres.2012.09.024>
- DABHOLKAR, P.A., THORPE, D.I., RENTZ, J.O. (1996). A measure of service quality for retail stores: Scale development and validation. **Journal of the Academy of Marketing Science**, Vol. 24 (1), pp. 3-16.
- DOĞAN, O., AYÇİN, E., & BULUT, Z. (2018). Customer segmentation by using RFM model and clustering methods: a case study in retail industry. **International Journal of Contemporary Economics and Administrative Sciences**, 8(1), (pp 1- 19).

- DURSUN, A., & CABER, M. (2016). Using data mining techniques for profiling profitable hotel customers: An application of RFM analysis. **Tourism management perspectives**, 18, 153-160.
- ERPOLAT TAŞABAT, S., & AKCA, E. (2020). Recycling Project with RFM Analysis in Industrial Material Sector. **Sigma Journal of Engineering and Natural Sciences**, 38(4), 1681-1692.
- F. A. BACHTIAR, "Customer Segmentation Using Two-Step Mining Method Based on RFM Model," 2018 International Conference on Sustainable Information Engineering and Technology (SIET), **Malang, Indonesia**, 2018, pp. 10-15, doi: 10.1109/SIET.2018.8693173.
- F. YOSEPH AND M. HEIKKILÄ,(2018) "Segmenting Retail Customers with an Enhanced RFM and a Hybrid Regression/Clustering Method," 2018 International Conference on Machine Learning and Data Engineering (iCMLDE), Sydney, NSW, Australia, 2018, pp. 108-116, doi: 10.1109/iCMLDE.2018.00029.
- GONZÁLEZ MARTÍNEZ, CARRASCO, GARCÍA-MADARIAGA, PORCEL GALLEGO, HERRERA-VIEDMA, (2019) "A comparison between Fuzzy Linguistic RFM Model and traditional RFM model applied to Campaign Management. Case study of retail business." **Procedia Computer Science Volume** 162, 2019, Pages 281-289 <https://doi.org/10.1016/j.procs.2019.11.286>
- HU, Y. H. & YEH, T. W. (2014). Discovering valuable frequent patterns based on RFM analysis without customer identification information. **Knowledge-Based Systems**, 61(3), 76-88
- JAMES, G., WITTEN, D., HASTIE, T., & TIBSHIRANI, R. (2013). An Introduction to Statistical Learning: with Applications in R (**Springer Texts in Statistics**) (1st ed. 2013, Corr. 7th printing 2017 ed.). Springer.
- KHAJVAND, M., ZOLFAGHAR, K., ASHOORÌ, S., & ALIZADEH, S. (2011). Estimating customer lifetime value based on RFM analysis of customer purchase behavior: Case study. **Procedia Computer Science**, 3, 57–63. <https://doi.org/10.1016/j.procs.2010.12.011>

- KİM, D., LEE, J. Y., AHN, S., MOON, Y., & KWON, O. J. (2012). RFM analysis for detecting future core technology. Proceedings of the 2012 ACM Research in Applied Computation Symposium on- **RACS'12**. <https://doi.org/10.1145/2401603.240161>
- KÖSE, U., & ARSLAN, A. (2020). A Novel Customer Segmentation Approach Based on RFM and Clustering: A Case Study in the Retail Industry. Gaziantep University **Journal of Social Sciences**, 19(4), 1229-1248. <https://doi.org/10.17671/gazibtd.570866>
- LANTZ, B. (2013). **Machine Learning with R**. Packt Publishing.
- MATUSZELAŃSKI, K., & KOPCZEWSKA, K. (2022). Customer Churn in Retail E-Commerce Business: Spatial and Machine Learning Approach. **Journal of Theoretical and Applied Electronic Commerce Research**, 17(1), 171-195. <https://doi.org/10.3390/jtaer17010009>
- MCCARTY, J., & HASTAK, M. (2007). Segmentation approaches in data-mining: A comparison of RFM, CHAID, and logistic regression. **Journal of Business Research**, (pp. 656-662).
- MİGLAUTSCH, J. R. (2000). Thoughts on RFM scoring. **Journal of Database Marketing & Customer Strategy Management**, 8(1), 67-72.
- NIKUMANESH, E., & ALBADVİ, A. (2014). Customer's life-time value using the RFM model in the banking industry: a case study. **International Journal of Electronic Customer Relationship Management**, 8(1/2/3), 15. <https://doi.org/10.1504/ijecrm.2014.066876>
- PAKYUREK, M., SEZGİN, M. S., KESTEPE, S., BORA, B., DUZAGAC, R., & YILDIZ, O. T. (2018). Customer clustering using RFM analysis. 2018 26th **Signal Processing and Communications Applications Conference (SIU)**. <https://doi.org/10.1109/siu.2018.8404680>
- PARİKİH, Y., & ABDELFAH, E. (2020). Clustering Algorithms and RFM Analysis Performed on Retail Transactions. 2020 11th IEEE Annual Ubiquitous Computing, Electronics & Mobile Communication

<https://doi.org/10.1109/uemcon51285.2020.9298123>

- PEKER, S., KOÇYIĞIT, A., & EREN, P. E. (2017). LRFMP model for customer segmentation in the grocery retail industry: a case study. *MARKETING INTELLIGENCE & PLANNING*, 544–559. <https://hdl.handle.net/11511/31802>
- PURNOMO, MUHAMMAD & ANUGERAH, ADHE & AZZAM, ABDULLAH & KHASANAH, ANNISA & ALFAREZA, M. (2021). RFM-based Customers Clustering for Precise **Industrial Marketing Strategy Formulation**. 15
- REICHHELD, F. AND SASSER, W. (1990). Zero defections: quality comes to services. *Harvard Business Review*, Vol. 68 (5), pp. 105-11.
- SARVARI, P. A., USTUNDAG, A., & TAKCI, H. (2016). Performance evaluation of different customer segmentation approaches based on RFM and demographics analysis. *Kybernetes*, 45(7), 1129–1157. <https://doi.org/10.1108/k-07-2015-0180>
- SHCHEKOLDIN, V. Y., & TSOY, M. Y. (2018). The Application of Modified RFM Analysis to Increase the Loyalty of Consumers of Industrial Rubber Articles. 2018 XIV International Scientific-Technical Conference on Actual Problems of Electronics Instrument Engineering (APEIE). <https://doi.org/10.1109/apeie.2018.8546229>
- STARCZEWSKI, A., & KRZYŻAK, A. (2015). Performance Evaluation of the Silhouette Index. *Artificial Intelligence & Soft Computing: 14th International Conference, ICAISC 2015, Zakopane, Poland, June 14-28, 2015, Proceedings, Part II*, 49–58. https://doi.org/10.1007/978-3-319-19369-4_5
- TAŞABAT, S. E., ÖZÇAY, T., SERTBAŞ, S., & AKCA, E. (2023). A New RFM Model Approach: RFMS. In *Industry 4.0 and the Digital Transformation of International Business* (pp. 143-172). **Singapore: Springer Nature Singapore.**

- WAN, S., CHEN, J., QI, Z., GAN, W., & TANG, L. (2022, April). Fast RFM model for customer segmentation. In Companion Proceedings of the **Web Conference 2022** (pp. 965-972).
- WEI, JO-TING & LIN, SHIH-YEN & WU, HSIN-HUNG. (2010). A review of the application of RFM model. **African Journal of Business Management December Special Review**. 4. 4199-4206
- YEH, I. C., YANG, K. J., & TING, T. M. (2009). Knowledge discovery on RFM model using Bernoulli sequence. **Expert Systems with Applications**, 36(3), 5866-5871
- YOU, Z., SI, Y. W., ZHANG, D., ZENG, X., LEUNG, S. C., & LI, T. (2015). A decisionmaking framework for precision marketing. **Expert Systems with Applications**, 42(7), 3357–3367. <https://doi.org/10.1016/j.eswa.2014.12.022>
- ZHANG, J., & SHEN, L. (2014). An Improved Fuzzy c-Means Clustering Algorithm Based on Shadowed Sets and PSO. ***Computational Intelligence and Neuroscience**, 2014*, Article ID 368628.

ELEKTRONİK KAYNAKLAR

- JAMES, G., WITTEN, D., HASTIE, T., & TIBSHIRANI, R. (2013). An introduction to statistical learning (Vol. 112, p. 18). New York: springer. Retrieved from <https://link.springer.com/content/pdf/10.1007/978-1-4614-7138-7.pdf> %20http://link.springer.com/10.1007/978-1-4614-7138-7.pdf
- YANG, A. X. (2004). How to develop new approaches to RFM segmentation. **Journal of Targeting, Measurement and Analysis for Marketing**, 13(1), 50–60. <https://link.springer.com/content/pdf/10.1057/palgrave.jt.5740131.pdf>

EKLER

Veri ön işleme kapsamında geliştirilen Python kodları aşağıda paylaşılmıştır. Bu kodlar preprocessing.py olarak adlandırılan proje dosyasından alınmıştır ve kodlara ilişkin yorumlar ve açıklamalar her bir kod satırı üzerinde bulunmaktadır.

```
import pandas as pd
import numpy as np
import constants

# CPI (Tüketici Fiyat Endeksi) verisini alır ve işler.
def get_cpi():
# Veriyi oku
cpi_data = pd.read_csv('TR_cpi.csv')

# Tarih formatını ayarla
cpi_data['date'] = pd.to_datetime(cpi_data['month_year'], format = '%m-%Y')

# Belirli bir tarih aralığına göre veriyi filtrele
cpi_data = cpi_data[(cpi_data.date > '2020-12-31') & (cpi_data.date < '2023-03-12')]

# CPI değerini kullanarak bir çarpan oluştur
cpi_data['multiplier'] = 1+cpi_data['cpi_mtm']/100

# Enflasyon çarpanını hesapla
cpi_data['inf_multiplier'] = cpi_data.multiplier.rolling(window=100, min_periods=1).agg(lambda x : x.prod())

return cpi_data

# Müşterilere göre fatura sayılarına dayanarak tek ve iki kez alışveriş yapan müşterileri alır.
def get_customers(df):
rec_count = df.groupby('Müşteri Kodu').agg({'Fatura Ref Numarası':'nunique'}).reset_index()
one_timers = rec_count[rec_count['Fatura Ref Numarası']==1]['Müşteri Kodu'].reset_index(drop=True)
two_timers = rec_count[rec_count['Fatura Ref Numarası']==2]['Müşteri Kodu'].reset_index(drop=True)
return one_timers.to_list(), two_timers.to_list()
```

```

# Müşteri başına alışveriş aralıklarının istatistiklerini alır.
def get_interval_stats(df):
    df = df.sort_values(['Müşteri Kodu', 'Fatura Tarihi'])
    df['Fatura Tarihi+_1'] = df.groupby('Müşteri Kodu')['Fatura Tarihi'].shift()
    df['interval'] = (df['Fatura Tarihi'] - df['Fatura Tarihi_1']).dt.days
    df = df.groupby('Müşteri Kodu').agg(interval_mean=('interval', 'mean'),
interval_std=('interval', 'std')).reset_index()
    return df

# Müşteri başına indirim oranını hesaplar.
def get_discount_ratio(df):
    df = df.groupby(['Müşteri Kodu', 'Fatura Ref Numarası']).agg({'Toplam
İsk': 'sum', 'UH_ANA GRUP': 'nunique'}).reset_index()
    df['is_discount'] = df['Toplam İsk'] > 0
    df = df.groupby('Müşteri Kodu').agg(order_count=('Fatura Ref Numarası',
'nunique'), discounted_order_count=('is_discount', 'sum'), depth=('UH_ANA
GRUP', 'mean')).reset_index()
    df['promo_ratio'] = df['discounted_order_count']/df['order_count']
    return df[['Müşteri Kodu', 'promo_ratio', 'discounted_order_count', 'depth']]

# Müşteri başına özellikler oluşturur.
def create_features(df):
    grouped_sample_data = df.groupby('Müşteri Kodu').agg(
receipt_date_max=('Fatura Tarihi', 'max'),
receipt_date_min=('Fatura Tarihi', 'min'),
monetary=(constants.sales_price_total, 'sum'),
order_count=('Fatura Ref Numarası', 'nunique')
)
    grouped_sample_data.reset_index(inplace=True)
    grouped_sample_data['receipt_date_max'] =
pd.to_datetime(grouped_sample_data['receipt_date_max'])
    grouped_sample_data['receipt_date_min'] =
pd.to_datetime(grouped_sample_data['receipt_date_min'])
    grouped_sample_data['time_gap'] =
grouped_sample_data['receipt_date_max']-
grouped_sample_data['receipt_date_min']
    grouped_sample_data_oc = grouped_sample_data.reset_index(drop=True)
    grouped_sample_data_oc['frequency'] =
grouped_sample_data_oc['order_count']/(grouped_sample_data_oc.time_gap/np.ti
medelta64(1, 'D'))
    grouped_sample_data_oc['recency'] =
(grouped_sample_data_oc['receipt_date_max'].max() -
grouped_sample_data_oc['receipt_date_max'])/np.timedelta64(1, 'D')
    grouped_sample_data_oc['length'] =
grouped_sample_data_oc.time_gap/np.timedelta64(1, 'D')
    return grouped_sample_data_oc

# Veri seti üzerinde özellik mühendisliği yapar.
def feature_engineering(df):

```

```

df['Fatura Tarihi'] = pd.to_datetime(df['Fatura Tarihi'])
one_timers, two_timers = get_customers(df)
cond = ((~df['Müşteri Kodu'].isin(one_timers)) & (~df['Müşteri
Kodu'].isin(two_timers)))
interval_data = get_interval_stats(df[cond])
discount_data = get_discount_ratio(df)
rest_features_data = create_features(df)
features = rest_features_data.merge(discount_data, on='Müşteri Kodu',
how='left').merge(interval_data, on='Müşteri Kodu', how='left')
features['one_timer'] = features['Müşteri Kodu'].isin(one_timers)
features['two_timer'] = features['Müşteri Kodu'].isin(two_timers)
return features

```

Kümeleme ile ilgili yapılan işlemlere ait kodlar aşağıda verilmiştir. Bu kodlar clustering.py dosyasından alınmıştır.

```
import pandas as pd
import numpy as np
import os
from sklearn.cluster import KMeans
from sklearn.metrics import silhouette_score
from sklearn.preprocessing import StandardScaler
from sklearn.metrics.pairwise import euclidean_distances
from sklearn import metrics
from scipy.spatial.distance import cdist
import skfuzzy as fuzz

def get_within_cluster(X, cluster_centers):
    # Hesapla ve döndür: Her veri noktası için en yakın küme merkezine olan
    ortalama uzaklık
    distortion = (sum(np.min(cdist(X, cluster_centers, 'euclidean'), axis=1)) /
X.shape[0])
    return distortion

def clustering_group_with_k(X, clustering_obj):
    # Veriyi ölçeklendir: Ortalaması 0 ve standart sapması 1 olacak şekilde
    scaler = StandardScaler()
    X_scaled = scaler.fit_transform(X)

    labels = {}
    sh_scores = {}
    inertias = {}
    distortions = {}

    # 5'ten 12'ye kadar olan k değerleri için kümeleme yap
    for n_clust in range(5,12):
        # Kümeleme modelini başlat
        co = clustering_obj(n_clusters=n_clust, max_iter=1000, n_init=5)
        co.fit(X_scaled)

        # Küme etiketlerini sakla
        labels[f'clusters_{n_clust}'] = co.labels_

        # Silhouette skorunu hesapla ve sakla
        sh_score = silhouette_score(X_scaled, co.labels_)
        sh_scores[f'clusters_{n_clust}'] = sh_score

        # Distortion değerini hesapla ve sakla
        distortions[f'clusters_{n_clust}'] = get_within_cluster(X_scaled,
co.cluster_centers_)

        # Inertia değerini sakla
        inertias[f'clusters_{n_clust}'] = co.inertia_
```

```

return [labels, sh_scores, distortions, inertias]

def clustering_fcmeans(X, **params):
    labels = {}
    sh_scores = {}
    distortions = {}

    # Veriyi ölçeklendir: Ortalaması 0 ve standart sapması 1 olacak şekilde
    scaler = StandardScaler()
    X_scaled = scaler.fit_transform(X)
    X_scaled_T = X_scaled.T

    # 5'ten 12'ye kadar olan k değerleri için fuzzy c-means kümeleme yap
    for n_clust in range(5,12):
        # Fuzzy c-means kümeleme modelini başlat ve fit et
        cntr, u, u0, d, jm, p, fpc = fuzz.cluster.cmeans(X_scaled_T, n_clust, m=2,
error=0.005, maxiter=1000)

        # Küme etiketlerini sakla
        label_list = np.argmax(u, axis=0)
        labels[f'clusters_{n_clust}'] = label_list

        # Silhouette skorunu hesapla ve sakla
        sh_score = silhouette_score(X_scaled, label_list)
        sh_scores[f'clusters_{n_clust}'] = sh_score

        # Distortion değerini hesapla ve sakla
        distortions[f'clusters_{n_clust}'] = get_within_cluster(X_scaled, cntr)

    return [labels, sh_scores, distortions]

```


Tüm projenin çalıştığı main.py dosyasına ait kodlar da aşağıdaki gibidir.

```
import pandas as pd
import numpy as np
import os
from preprocessing import get_interval_stats, get_discount_ratio,
create_features, feature_engineering, get_cpi
from clustering import clustering_group_with_k, clustering_fcmeans
import constants
from sklearn.cluster import KMeans
from sklearn.preprocessing import StandardScaler
from sklearn.metrics import silhouette_score
import skfuzzy as fuzz
import os
import matplotlib.pyplot as plt
import seaborn as sns
from scipy import stats
import pickle
from mpl_toolkits.mplot3d import Axes3D

def run():
    #Satış verisini okuma
    dfs_list = []
    for csv_file in os.listdir('Data'):
        dfs_list.append(pd.read_csv(f'Data/{csv_file}'))
    #Ayrık okunan satış verilerini birleştirme
    main_df = pd.concat(dfs_list, ignore_index=True)
    #Aylık enflasyon verisini alma
    cpi_df = get_cpi()
    #Fatura Tarihi verisini datetime tipine çevirme
    main_df['Fatura Tarihi'] = pd.to_datetime(main_df['Fatura Tarihi'])
    #Satış tutarı 0 ve 0'dan küçük olan satırları eleme
    main_df = main_df[main_df['Satış (VD)']>0].reset_index(drop=True)
    #Enflasyona bağlı fiyat düzenlemesi için ay ve yıl bilgisi çıkarma
    main_df['month_year'] = main_df['Fatura Tarihi'].dt.strftime('%m-%Y')
    #Enflasyon bilgisini veriye ekleme
    main_df = main_df.merge(cpi_data[['month_year', 'inf_multiplier']],
on='month_year', how='left')
    #Hesaplanan enflasyon ile satış tutarının son aya uyarlanması
    main_df.inf_multiplier = main_df.inf_multiplier.fillna(1)
    main_df['today_price'] = main_df['Satış (VD)']*main_df['inf_multiplier']

    #Feature engineering kodunun çağırılması
    output = feature_engineering(main_df)
    X = output[['Müşteri Kodu', 'monetary', 'order_count', 'recency',
'promo_ratio', 'depth', 'interval_std', 'length']]
    #Bool array ile bir ve iki siparişi olan müşterileri veriden ayıklama
    cond = (output.one_timer==False)&(output.two_timer==False)
    X_3 = output[cond].reset_index(drop=True)
```

```
output[['receipt_date_max', 'receipt_date_min', 'monetary', 'order_count',  
'length', 'frequency', 'recency', 'interval_mean',  
'depth']].describe().to_excel('data_desc.xlsx')
```

```
#Monetary değişkeni için dağılım görselleri oluturma ve dağılım dönüşümü  
yapma
```

```
X_3.monetary.plot(kind='hist')  
plt.show()  
fitted_data, fitted_lambda = stats.boxcox(X_3.monetary)  
sns.distplot(fitted_data, hist = True, kde = True,  
kde_kws = {'shade': True, 'linewidth': 2}, label = "Normal", color = "green")  
plt.show()  
X_3['monetary_transformed'] = fitted_data
```

```
#Monetary değişkeni için dağılım görselleri oluturma
```

```
sns.set(font_scale=2)  
fig, ax = plt.subplots(figsize=(15, 10))  
sns.histplot(  
data=output[output.monetary < 20000],  
x='monetary',  
stat='percent',  
bins=50,  
ax=ax  
)  
plt.savefig('monetary_dagilim.png', dpi = 1000)  
plt.plot()
```

```
#Sipariş sayısı değişkeni için dağılım görselleri oluturma ve dağılım  
dönüşümü yapma
```

```
X_3.query('order_count < 100').order_count.plot(kind='hist')  
plt.show()  
X_3.order_count.apply(np.log10).plot(kind='hist')  
plt.show()
```

```
fig, ax = plt.subplots(figsize=(15, 10))  
sns.histplot(  
data=output[output.order_count < 10],  
x='order_count',  
discrete=True,  
stat='percent',  
ax=ax  
)  
plt.savefig('order_count.png')  
plt.plot()
```

```
fitted_data, fitted_lambda = stats.boxcox(X_3.order_count)  
sns.distplot(fitted_data, hist = True, kde = True,  
kde_kws = {'shade': True, 'linewidth': 2}, label = "Normal", color = "green")  
X_3['order_count_transformed'] = fitted_data
```

```

#Recency değişkeni için dağılım görselleri oluturma ve dağılım dönüşümü
yapma
X_3.recency.plot(kind='hist', bins=50, figsize=(15,10))
fig, ax = plt.subplots(figsize=(15, 10))
sns.histplot(
    data=output,
    x='recency',
    stat='percent',
    bins=40,
    ax=ax
)
plt.savefig('recency.png')
plt.plot()
fitted_data, fitted_lambda = stats.boxcox(X_3.recency+0.00001)
sns.distplot(fitted_data, hist = True, kde = True,
kde_kws = {'shade': True, 'linewidth': 2}, label = "Normal", color = "green")
y = (400**fitted_lambda - 1) / fitted_lambda
print(y)
X_3['recency_transformed'] = fitted_data
X_3.loc[X_3.recency_transformed>=y, 'recency_transformed'] = y

#K-Means ve Fuzzy-c-means yöntemlerinin çalıştırılması
X_to_cluster = X_3[['monetary_transformed', 'order_count_transformed',
'recency_transformed']]
kmeans_rfm = clustering_group_with_k(X_to_cluster, KMeans)
X_to_cluster = X_3[['monetary_transformed', 'order_count_transformed',
'recency_transformed']]
fcmeans_rfm = clustering_fcmeans(X_to_cluster)

#Modellerden çıkan küme bilgileri, Silhoutte skorları ve küme için toplam
uzaklık skorlarının kaydedilmesi
with open('Cluster_results/kmeans_labels_rfm_new.pkl', 'wb') as f:
    pickle.dump(kmeans_rfm[0], f)
with open('Cluster_results/kmeans_sh_scores_rfm_new.pkl', 'wb') as f:
    pickle.dump(kmeans_rfm[1], f)
with open('Cluster_results/kmeans_distortion_scores_rfm_new.pkl', 'wb') as
f:
    pickle.dump(kmeans_rfm[2], f)
with open('Cluster_results/kmeans_inertia_scores_rfm_new.pkl', 'wb') as f:
    pickle.dump(kmeans_rfm[3], f)
with open('Cluster_results/fcmeans_labels_rfm_new.pkl', 'wb') as f:
    pickle.dump(fcmeans_rfm[0], f)
with open('Cluster_results/fcmeans_sh_scores_rfm_new.pkl', 'wb') as f:
    pickle.dump(fcmeans_rfm[1], f)
with open('Cluster_results/fcmeans_distortion_scores_rfm_new.pkl', 'wb')
as f:
    pickle.dump(fcmeans_rfm[2], f)

```

```

#Kümelemede kullanılan verinin okunması
data = pd.read_csv('main_data.csv')
#Modellerden çıkan küme bilgileri, Silhoutte skorları ve küme için toplam
uzaklık skorlarının okunması
gen_dict = {}
for file in os.listdir('Cluster_results'):
print(file)
if file.startswith('data') or file.startswith('main'):
continue
with open("Cluster_results/"+file, 'rb') as f:
file_dict = pickle.load(f)
gen_dict[file[:-4]] = file_dict

#Modellerden çıkan küme bilgileri, Silhoutte skorları ve küme için toplam
uzaklık skorlarının ayrı ayrı arraylere alınması
labels = []
distortions = []
sh_scores = []
for key in gen_dict.keys():
if 'labels' in key and 'new' in key:
labels.append(key)
elif 'distortion' in key and 'new' in key:
distortions.append(key)
elif 'sh_scores' in key and 'new' in key:
sh_scores.append(key)

#Silhouette skorlarının karşılaştırılması
for key in sh_scores:
print("SH SCORES FOR", key)
clustering_label_dict = gen_dict[key]
print(clustering_label_dict)

#Küme işaretlerinin genel veriye eklenmesi
for key in labels:
clustering_label_dict = gen_dict[key]
for c_key in clustering_label_dict.keys():
data[key+"_"+c_key] = clustering_label_dict[c_key]

#Küme merkezlerinin hesaplanması
label_cols = data.columns[data.columns.str.contains('labels')]
ccenter_dict = {}
for col in label_cols:
ccenter_df = data.groupby(col).mean().iloc[:, :-
len(label_cols)].sort_values(['monetary', 'recency'])
ranked_ccenter_df = ccenter_df.rank()
ranked_ccenter_df = ranked_ccenter_df.add_suffix('_rank')
ccenter_df=pd.concat([ccenter_df, ranked_ccenter_df], axis=1)
ccenter_df = ccenter_df.reindex(sorted(ccenter_df.columns), axis=1)
ccenter_df = pd.concat([ccenter_df, data.groupby(col).size()], axis=1)
ccenter_dict[col] = ccenter_df.rename(columns={0:'size'})

```

```

#WCSS skorlarının karşılaştırılması
for key in distortions:
    clustering_label_dict = gen_dict[key]
    c_num = list(clustering_label_dict.keys())
    vals = list(clustering_label_dict.values())
    plt.plot(c_num, vals)
    plt.title(key)
    plt.savefig(key+'.png', dpi=800)
    plt.show()

#Küme merkezlerinin kaydedilmesi
ccenter_dict['fcmeans_labels_extended_rfm_new_clusters_5'].to_excel('fcmeans5.xlsx')
ccenter_dict['kmeans_labels_extended_rfm_new_clusters_8'].to_excel('kmeans8.xlsx')

#RFM değerlerinin gösterimi fuzzy c-means için
X_to_cluster = data[['monetary_transformed', 'order_count_transformed', 'recency_transformed']]
fig, ax = plt.subplots(3,3, figsize=(40, 20))
for i in range(3):
    for j in range(3):
        ax[i,j].scatter(X_to_cluster.iloc[:, i], X_to_cluster.iloc[:, j],
c=data['fcmeans_labels_rfm_new_clusters_5'], s=50, cmap='viridis')
        ax[i,j].set_xlabel(X_to_cluster.columns[i])
        ax[i,j].set_ylabel(X_to_cluster.columns[j])
    plt.savefig('fcmenas5_all_features.png')
    plt.plot()

#RFM değerlerinin gösterimi K-means için
X_to_cluster = data[['monetary_transformed', 'order_count_transformed', 'recency_transformed']]
fig, ax = plt.subplots(3,3, figsize=(40, 20))
for i in range(3):
    for j in range(3):
        ax[i,j].scatter(X_to_cluster.iloc[:, i], X_to_cluster.iloc[:, j],
c=data['kmeans_labels_rfm_new_clusters_8'], s=50, cmap='viridis')
        ax[i,j].set_xlabel(X_to_cluster.columns[i])
        ax[i,j].set_ylabel(X_to_cluster.columns[j])
    plt.savefig('kmenas8_all_features.png')
    plt.plot()

#K-Means dönüştürülmüş değişkenlerle 3 boyutlu görselleştirme
fig = plt.figure(figsize=(20,20))
ax = fig.gca(projection='3d')
xs = X_to_cluster['monetary_transformed']
ys = X_to_cluster['recency_transformed']
zs = X_to_cluster['order_count_transformed']
ax.scatter(xs, ys, zs, c=data['kmeans_labels_rfm_new_clusters_8'])
ax.set_xlabel('monetary_transformed')

```

```
ax.set_ylabel('recency_transformed')
ax.set_zlabel('frequency_transformed')
plt.savefig('3d_kmeans_8_t.png', dpi = 1000)
plt.show()
```

```
#Fuzzy C-Means dönüştürülmüş değişkenlerle 3 boyutlu görselleştirme
fig = plt.figure(figsize=(20,20))
ax = fig.gca(projection='3d')
xs = X_to_cluster['monetary_transformed']
ys = X_to_cluster['recency_transformed']
zs = X_to_cluster['order_count_transformed']
ax.scatter(xs, ys, zs, c=data['fcmeans_labels_rfm_new_clusters_5'])
ax.set_xlabel('monetary_transformed')
ax.set_ylabel('recency_transformed')
ax.set_zlabel('frequency_transformed')
plt.savefig('3d_fcmeans_5_t.png', dpi = 1000)
plt.show()
```

```
#K-Means orijinal değişkenlerle 3 boyutlu görselleştirme
cond =( data.order_count<20) & (data.monetary<15000)
X_to_cluster = data[cond]
fig = plt.figure(figsize=(20,20))
ax = fig.gca(projection='3d')
xs = X_to_cluster['monetary']
ys = X_to_cluster['recency']
zs = X_to_cluster['order_count']
ax.scatter(xs, ys, zs, c=X_to_cluster['kmeans_labels_rfm_new_clusters_8'])
ax.set_xlabel('monetary')
ax.set_ylabel('recency')
ax.set_zlabel('frequency')
plt.savefig('3d_kmeans_8_orig.png', dpi = 1000)
plt.show()
```

```
#K-Means dönüştürülmüş değişkenlerle 3 boyutlu görselleştirme
cond =( data.order_count<20) & (data.monetary<15000)
X_to_cluster = data[cond]
fig = plt.figure(figsize=(20,20))
ax = fig.gca(projection='3d')
xs = X_to_cluster['monetary']
ys = X_to_cluster['recency']
zs = X_to_cluster['order_count']
ax.scatter(xs, ys, zs, c=X_to_cluster['fcmeans_labels_rfm_new_clusters_5'])
ax.set_xlabel('monetary_transformed')
ax.set_ylabel('recency_transformed')
ax.set_zlabel('frequency_transformed')
plt.savefig('3d_fcmeans_5_orig.png', dpi = 1000)
plt.show()
```

```
if __name__ == '__main__':
    run()
```

ÖZGEÇMİŞ

KİŞİSEL BİLGİLER

Adı-Soyadı: Samet Kanca

EĞİTİM

Derece	Kurum	Mezuniyet
Önlisans	Beykent Üniversitesi	2008 – 2010
Lisans	Anadolu Üniversitesi	2010 – 2015
Yüksek Lisans	İstanbul Aydın Üniversitesi	2021 – Devam ediyor

İŞ DENEYİMİ

Yıl	Kurum	Görev
2010 – 2012	Getron Yazılım	Yazılım Geliştirme Uzmanı
2012 – 2018	Getron Yazılım	Ürün Müdürü
2018 – 2021	Getron Yazılım	İş Geliştirme Direktörü
2021 – 2023	Necad Yazılım	Teknik İş Geliştirme Lideri
2023 –	Saran Holding	Süreç Geliştirme Direktörü