

Müşteri Odaklı Pazarlama Stratejileri için Veri Madenciliği Teknikleri Kapsamında Perakende Sektöründe Kümeleme Analizi Uygulaması¹

Application of Clustering Analysis in Retail Sector within the Scope of Data Mining Techniques for Customer Oriented Marketing Strategies

Ersin Namlı², Sümeyra Murat³

Öz

Günümüz rekabetçi ortamında pazarlama kavramının sadece satış ve reklam anlayışı ile sınırlanmayacağı ekonomik göstergeler ve pazar paylarındaki değişimler neticesinde kolaylıkla anlaşılmaktadır. Pazarlamada ki temel noktanın tüketici ihtiyaçlarının araştırılıp ortaya çıkarılmasından, satın alma sonrası müşteri sadakati ve memnuniyetine kadar olan geniş bir süreç olduğu yapılan bilimsel çalışmalar ve araştırmalarla ortaya konmuştur.

Tüketicilerin ürün veya hizmet satın alırken nasıl karar verdikleri kişisel, sosyal ve psikolojik olarak hangi parametrelerden etkilendikleri satış sonrası firmalardan beklentileri tüketicilerin davranışlarını belirlemektedir. İşletmeler pazarlama stratejilerini tüketicileri çeşitli gruplara ayırarak, davranışları benzerlik gösteren müşterilerine benzer stratejileri uygulayarak Pazar paylarını, marka değerlerini ve karlılıklarını arttırmaya çalışmaktadırlar. Buna ek olarak Firmalar müşteri ilişkilerini verimli ve faydalı bir biçimde yönetebilmek için birçok bilimsel yöntem kullanarak müşteri segmentasyonu yapmaktadırlar. Son yıllarda popülaritesi gittikçe artan veri madenciliği tekniklerinden, sınıflandırma, kümeleme ve birliktelik analizi gibi yöntemlerin oldukça başarılı sonuçlar verdiği görülmektedir.

Bu çalışmada ülkemizin lokomotif sektörlerinden olan tekstil sektörü kapsamında perakende tekstil satış verileri incelenmiştir. İki Aşamalı Kümeleme Analizi yöntemi ve Beklenti Maksimizasyonu algoritması kullanılarak müşteri gruplarına özel promosyon ve kampanya çalışmalarına ek olarak müşteri odaklı pazarlama stratejileri geliştirmek için kümeleme analizi yapılmıştır. Elde edilen müşteri segmentasyon sonuçları doğrultusunda firma için akıllı pazarlama önerileri ve stratejiler ortaya konmuştur.

Anahtar Kelimeler: Veri Madenciliği, Kümeleme, Tüketici Davranışları, Pazarlama,

Abstract

In today's competitive environment, marketing concept can not be limited only by sales and advertising understanding, as a result of economic indicators and changes in market shares is easily understood. The main point in marketing is a wide process from exploring and revealing consumer needs, post-purchase service, customer loyalty and satisfaction. is laid out by studies and research

How consumers decide when buying products or services, which parameters are affected by personal, socially and psychologically, the expectations of after-sales firms are consumers' determines their behavior. Business marketing strategies by dividing consumers into a variety of groups, applying similar strategies to their customers whose behavior is similar to their customers, market shares, brand values and are trying to increase their profitability. In addition, companies are engaged in customer segmentation using many scientific methods to manage customer relationships efficiently and beneficially. From the increasingly popular data mining techniques in recent years, methods such as classification, clustering and association analysis show quite successful results.

In this study, retail textile sales data were analyzed within the scope of the textile sector, which is one of the locomotive sectors of our country. Customer-oriented marketing strategies in addition to promotional and campaign work for customer groups using Two-Step Clustering Analysis method and Expectation Maximization algorithm was conducted clustering analysis to improve. Intelligent marketing proposals and strategies for the company are presented in accordance with the results of the customer segmentation.

Keywords: Data Mining, Clustering, Customer Behavior, Marketing.

Araştırma Makalesi [Research Paper]

JEL: L1, L3

Submitted: 16 / 09 / 2019

Accepted: 13 / 10 / 2019

¹ Bu makalenin özeti 18-20 Nisan 2019 tarihinde gerçekleştirilen "International Scientific Researches Congress 2019 Tbilisi" 'de bildiri olarak sunulmuştur.

² Istanbul University – Cerrahpasa, Industrial Engineering Department, Istanbul/Turkey

³ Istanbul University – Cerrahpasa, Industrial Engineering Department, Istanbul/Turkey

Giriş

Teknoloji, iletişim, taşımacılık ve ulaşımdaki hızlı gelişimin etkisiyle pazarlama kavramı firmalar için önemli bir konuma gelmiştir. Pazarlama birimi üretici ile son kullanıcı yani müşteri arasındaki iletişim ağını oluşturan yapıdır. Hangi müşterinin hangi ürüne ne kadar ödeyeceği, son kullanıcıların neye ihtiyacı olduğu, bir ürünün üretim maliyetinin ne kadar olmasına direk etki eden karar mekanizmasıdır. Bir firma üretimde ve tedarikte ne kadar başarılı olursa olsun eğer pazarlama birimi zayıf ve yetersiz ise pazarlama kavramı ve anlayışı benimsenememiş ise sadece başka bir marka için üretim yapabilen fason bir şirket olarak tanımlanmaktan kurtulamaz. Üretim ve lojistikteki gelişmeler ve yoğun rekabet firmalar bu alanlarda kendilerini geliştirmeye mecbur bırakmıştır üretim ve lojistikte belirli bir gelişmişlik seviyesini yakalayan işletmeler rekabette geri kalmamak hatta rakiplerinden daha ileriye gidebilmek için pazarlama kavramı üzerine yoğunlaşmışlardır. Özellikle kalite kavramının önem kazanmasıyla şirketlerin üretim odaklı yönetim anlayışından müşteri odaklı yönetim anlayışına geçmeleriyle bir başka değişle müşteri istek ve ihtiyaçları doğrultusunda bir yönetim anlayışı benimsenmesiyle modern pazarlama kavramı doğmuştur.

Modern pazarlamada işletme karlılığı anlayışı müşteri tatmini anlayışı çerçevesinde şekillenmektedir. Müşteri odaklı bir anlayışta öncelikle müşteriyi tanımakla başlar. İçinde bulunduğumuz yüzyılda veri yeni petrol olarak tanımlanmaktadır ve gelecekte verinin ekonomik katma değerinin petrolü geride bırakacağı düşüncesi kabul görmüştür. Veri bilimini kısaca verinin kayıt altına alınması, saklanması, işlenmesi ve faydalı enformasyon ve bilgiye dönüşümü için işlenmesi olarak tanımlayabiliriz. Modern pazarlamada veri bilimi müşterilerin davranışlarını, satın alma alışkanlıklarını, müşteri bildirimlerini analiz edilerek müşteri alışkanlıklarını çözümlenmesine olanak sağlamaktadır ve bunun sonucunda firmalar müşteri odaklı çözümler, kampanyalar ve satış stratejileri geliştirebilmektedirler. Yapılan internet aramaları, satın alınan ürünler, ziyaret edilen mağazalar, gidilen lokasyonlar, sosyal medya paylaşımları, ürün raflarında geçirilen süreler, renk ve beden tercihleri gibi birçok veri müşteri alışkanlıklarının analizi için değerli kaynaklar olarak gösterilebilir. Çalışmamızda pazarlama biriminin kaynaklarını ve bütçesini verimli bir biçimde yönetebilmek ve etkin stratejilerle firma gelirini artırmak amacıyla bir tekstil firmasının müşteri odaklı pazarlama ve kampanya stratejileri geliştirebilmesi için kümeleme analizi yapılmıştır ve müşteri satınalma davranışlarına göre pazarlama önerileri sunulmuştur.

1. Pazarlama ve Veri Madenciliği

Pazarlamayı tanımlayacak olursak, bazı işletme yöneticileri dahil olmak üzere çoğu insan, pazarlamanın satış veya reklam anlamına geldiğini söyler. Bunların pazarlamanın bir parçası olduğu doğrudur. Ancak pazarlama, satış ve reklamcılıktan çok daha fazlasıdır. Günümüzde modern pazarlama olarak tanımlanan kavram şirket karlılığı çerçevesinde müşteri odaklı stratejiler geliştirmek olarak açıklanabilir

Günümüzde veri bilimi kapsamında veri madenciliği tekniklerinin tüm iş alanlarında ve disiplinlerde yoğun bir biçimde kullanıldığı bu konuda yayınlanan çalışmaların artan bir ivmeye sahip olmasından anlaşılabilir. Özellikle dijital teknolojinin gelişmesi ve e-ticaretin yaygınlaşmasıyla veri madenciliği teknikleri pazarlama bilimi için ayrı bir konuma gelmiştir. Firmalar müşterilerini tanımak ve ilişkileri geliştirmek, mevcut müşterilerin alışveriş sıklıklarını artırmak, kayıp müşterilerin analizini yaparak onları tekrar kazanmak, müşterilerinin hangi periyotlarda ne kadar harcama yapacaklarını tahmin etmek. Müşterileri segmentasyonu gibi pazarlama stratejilerine altyapı oluşturacak birçok parametreyi veri madenciliği teknikleriyle analiz etmektedirler.

Erevelles ve diğ. (2016) çalışmalarında büyük veri analizinin firmalara sağladığı avantajları araştırmışlar ve veri biliminin firmaların pazarlama stratejilerindeki değişime olan etkilerini ortaya koymuşlardır. Chong ve diğ. (2017) ortaya koydukları çalışmada çevrim içi promosyonel pazarlamanın ve kullanıcılara ait çevrimiçi yorumların ürün taleplerini nasıl etkilediğini incelemişlerdir ve bahsi geçen faktörlerin üreticilerin hesaba katması gereken faktörler olduğunu ortaya koymuşlardır. Gordini ve Veglio (2017) yeni bir modelin tahmin kapasitesini (SVMauc) olarak tanımladıkları parametre seçimi ile destek vektör makinesi (DVM) kullanarak test etmişlerdir ve çalışmanın sonucunda B2B e-ticaret endüstrisi için uyarlanmış bir müşteri kaybı öngörü modeli geliştirmişlerdir. Bahari ve Elayidom (2015) Yapay Sinir Ağları (YSA) ve Naive Bayes (NB) tekniklerini kullanarak 45211 adet banka pazarlama kampanya verilerinde sınıflandırma çalışması yapmışlardır çalışmalarında YSA yönteminin %88,63'lük bir doğru sınıflandırma başarısıyla veri madenciliği ve makine öğrenme tekniklerinin müşterilerin sınıflandırılmasında etkin bir yöntem olduğunu ispatlamışlardır. Kumar ve diğ. (2019) pazarlama alanında yapay zekanın rolü konusunu ele alarak yapay zekanın müşteri yönetim stratejileri yönetimi üzerindeki etkilerinden bahsetmişlerdir. Çalışmalarında ayrıca gelişmiş ve gelişmekte olan ülkelerde pazarlama için yapay zeka tabanlı kişiselleştirilmiş sözleşme kavramını incelemişlerdir.

Müşteri davranışlarına veya özelliklerine göre pazarlama ve kampanya stratejileri geliştirmek için uygulanan müşterileri segmentasyonu sürecinde kümelere algoritmalarının oldukça yoğun kullanıldıkları görülmektedir. Geleneksel kümeleme algoritmalarının yetersiz kaldığı durumlarda konsensus kümeleme algoritmalarıyla daha net analiz çözümlerine

ulaşmaktadır (Ünlü ve Xanthopoulos, 2019a,b) Fuentes ve diğ. (2018) Müşteri satın alma davranışlarını çok örnekli kümeleme yaklaşımıyla analiz ederek güvenilir bir segmentasyon modeli geliştirmişlerdir. Claveria ve Poluzzi (2017) çalışmalarında hedef konumlandırma çalışmalarında kullanılacak bir model geliştirerek, dünyanın en önemli turizm bölgelerini 2000 ve 2010 yılları arasındaki ana turizm göstergeleri kapsamında kümelemişlerdir. Hol ve diğ. (2017) Eczacılık sektöründe piyasa sepeti verilerini kullanarak müşteri davranışına göre perakende ürünlerinin kümelmesi için bir yöntem önermişlerdir. Biscarri ve diğ (2017) kümeleme yöntemi ile otomatik sınıflandırma modeli geliştirerek müşteriler tarafından kullanılan elektrik yüklerini sınıflandırmışlardır. Bu sayede tedarikçi firmalar için elektrik kullanımlarına göre sınıflandırılan müşterilere göre pazarlama stratejisi geliştirme imkanı sağlamışlardır.

2. Veri Seti ve Metodoloji

Uygulamadan kullanılan veri seti bir tekstil firmasının müşterilerine özel kart tanımlaması sonucu elde ettiği müşteri bilgilerinden oluşmaktadır. Müşteri bilgilerinin birçoğu demografik özellik oluşturmasıyla birlikte, satış tutarından, kaç farklı ürün grubundan alışveriş yaptığı bilgileri gibi çeşitli verilerde elde edilmektedir.

Tablo 1. Veri Seti Özellikleri

Yas	Cinsiyet	FisSayisi	SatisAdedi	Ciro	İl	Ana Mağaza	Eğitim Durumu	Farklı Ürün Grubu Sayısı	Ağırlıklı Ürün Miktarı
39	Erkek	27	44	7.433 TL	Antalya	Fiziksel	Lisansüstü	15	Kampanyasız
42	Kadın	30	46	7.309 TL	Tekirdağ	Fiziksel	Lisans	16	Kampanyasız
38	Erkek	17	46	6.782 TL	Mersin	Fiziksel	ÖnLisans	16	Kampanyasız
30	Kadın	21	43	6.508 TL	Aydın	Fiziksel	Lisans	15	Kampanyalı
35	Erkek	27	43	6.245 TL	İstanbul	Fiziksel	ÖnLisans	15	Kampanyalı
44	Erkek	18	46	6.017 TL	İstanbul	Fiziksel	Lise	16	Kampanyalı
28	Erkek	21	33	5.923 TL	İstanbul	Fiziksel	ÖnLisans	11	Kampanyalı
44	Kadın	26	41	5.856 TL	Denizli	Fiziksel	Lisans	14	Kampanyalı

Veride bulundurulmuş müşteri bilgileri 11 başlık altında toplanmış olup içerik olarak şu şekilde ifade edilmektedir;

- Müşteri ID
- Yaşı
- Cinsiyeti
- Toplamda aldığı fiş sayısı
- Satın alınan ürün adedi
- Satın alınan tüm ürünlerden elde edilen ciro
- Bulunduğu il
- Ürünün satın alındığı platform
- Eğitim durumu
- Kaç farklı ürün grubundan alışveriş yaptığı
- Ağırlıklı olarak hangi ürün grubundan alışveriş yaptığı

Çalışmada toplam 20.181 veri kullanılmıştır. Veri önleme aşamasında veride eksik bilgi olmaması, tek bir kaynak üzerinden veri setinin oluşturulmuş olması ve uyumsuzluk oluşturacak bir değere rastlanmaması sonucu çıkarılmıştır.

Verinin hangi veri madenciliği algoritması ile daha iyi sonuçlar vereceği ve pazarlama stratejisi geliştirilebilecek sonuçlar çıkarmak için IBM SPSS Statistics yazılımı kullanılmıştır. Bu veri setindeki müşteri profilini analiz etmek ve yorumlamak, çıkan sonuçlara uygun müşteri beklentilerini karşılamak adına çalışmanın ilerleyen kısımlarında bahsedilecek olan kümeleme algoritmalarından yararlanılmıştır.

Veri setinde hem kategorik hem de nümerik veri bulunduğundan dolayı iki türde verinin aynı veri setinde bulunduğu durumlarda kullanılan "İki Adımlı Kümeleme Yöntemi" (A Two Step Method) kullanılmıştır. Çoğu kümeleme algoritması ya tam saf sayısal veriler ya da tam saf kategorik veriler üzerinde çalışabilir. İki adımlı kümeleme algoritması ise karma olan verilerin kümelendiğinde daha sağlam sonuçların elde edileceği deneysel kanıtlarla gösterilmektedir. (Shih ve diğ., 2010)

Küme sayısının algoritma tarafından belirlendiği İki Aşamalı kümeleme (Two Step Clustering) yöntemi, büyük veri kümelerinin analizi için Chiu ve diğ. (2001) tarafından geliştirilmiştir. Yöntem iki adımdan oluşmaktadır.

1. Adım: Olguların önceden kümelenmesi adıdır. Sıralı bir yaklaşım kullanılır. Amaç bir sonraki adım için daha az olgu ile yeni veri matrisi hesaplamaktır. Bu aşamada 2. Adım'da kullanılmak üzere ön kümeler belirlenir. Ön kümeleme yoğunlukta olan nitelikler için hesaplanarak belirlenir.

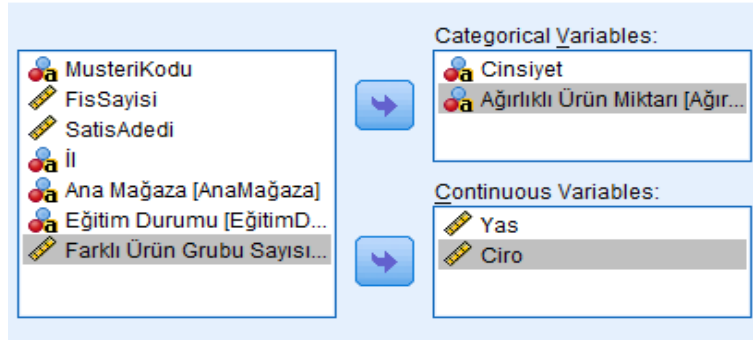
2. Adım: Belirlenen ön kümeler bu adımda model tabanlı hiyerarşik yöntem kullanılarak, birbirine en yakın kümeler her adımda birleştirilerek tüm kümeler birleşene kadar devam etmektedir.

İki Aşamalı kümeleme algoritması haricince "Beklenti Maksimizasyonu" (Expectation Maximization) kümeleme yöntemi de içerisinde kategorik ve nümerik nitelikler barındıran veri setlerini kümelemek için uygundur. Beklenti Maksimizasyonu algoritması karma model çerçevesi ile yoğunluk tahmini kümeleme problemlerinin genelleştirilmesi olarak görülebilir. Karma model parametrelerini hesaplamak için etkin bir yöntemdir. Diğer yinelemeli tekniklere göre istatistiksel bakış açısıyla modelleme yeteneği daha kuvvetlidir. (Bradley ve diğ., 1998)

3. Bulgular

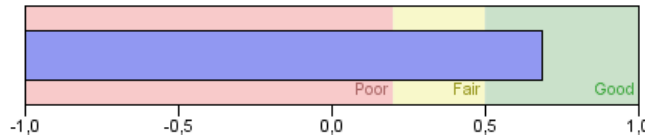
3.1 İki Aşamalı Kümeleme Analizi

İki Aşamalı (Two Step) algoritma ile yapılan denemeler sonucu en iyi ortalama silhoutte katsayısının elde edildiği nitelik seçimi şekil 1'deki gibidir;

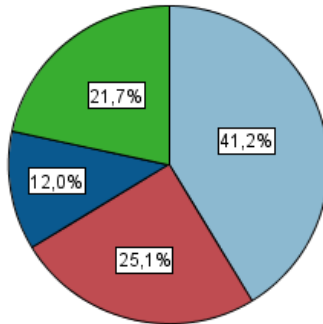


Şekil 1. İki Aşamalı Kümeleme Analizi Nitelik Seçimi

4 girdi değişkeninin kullanıldığı iki aşamalı kümeleme sonucunda silhoutte katsayısı 0,7 olan 4 farklı kümeden oluşan bir sonuç elde edilmiştir. Kümelerin dağılım yüzdeleri şekil 3'deki gibidir.







Şekil 2. Örnek 1 Silhouette katsayısına endeksli Kümeleme Kalitesi

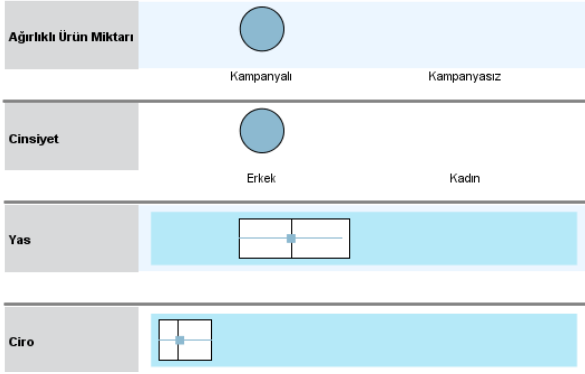


Şekil 3. Kümelerin Dağılım Yüzdeleri

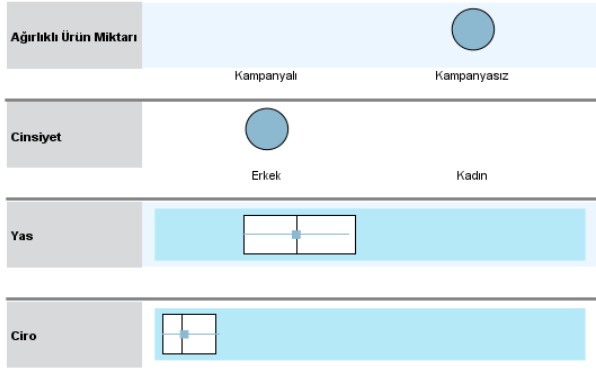
Tablo 2. Örnek 1 Küme İçi Dağılımlar

Size	Inputs			
 41,2% (5766)	Ağırlıklı Ürün Miktarı Kampanyalı (100,0%)	Cinsiyet Erkek (100,0%)	Yaş 36,43	Ciro 261,91
 25,1% (3512)	Ağırlıklı Ürün Miktarı Kampanyasız (100,0%)	Cinsiyet Erkek (100,0%)	Yaş 36,29	Ciro 268,39
 12,0% (1679)	Ağırlıklı Ürün Miktarı Kampanyasız (100,0%)	Cinsiyet Kadın (100,0%)	Yaş 37,35	Ciro 250,77
 21,7% (3030)	Ağırlıklı Ürün Miktarı Kampanyalı (100,0%)	Cinsiyet Kadın (100,0%)	Yaş 37,73	Ciro 257,50

Şekil 3'te kümelerin yüzdeler dağılım grafiği Tablo 2'de ise küme içi niteliklerin yüzdeler dağılımları ve detayları görülmektedir. %41.2 değerle 1. Kümeyi oluşturan müşterilerin hepsinin; cinsiyeti erkek olmakla birlikte, ağırlıklı olarak kampanyalı ürün tercih eden, yaş ortalaması 36,43 olan ve kazandırdıkları ciro ortalaması 261,91'dir bu durumda az da olsa veri seti ortalamasının üstünde ciro getirmiştir. 1. kümeye ait nitelikler şekil 4'te gösterilmiştir.



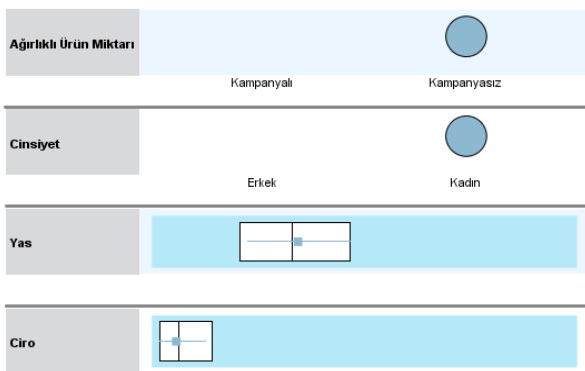
Şekil 4. Küme 1 Nitelikler



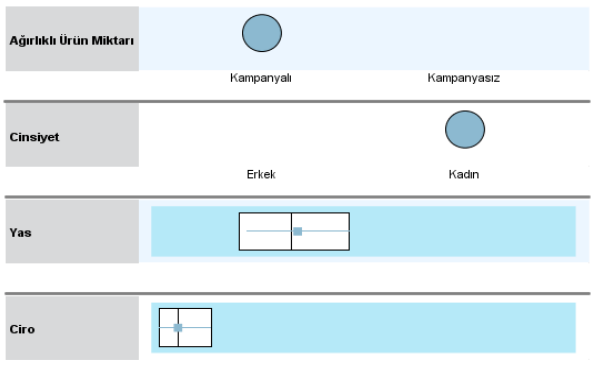
Şekil 5. Küme 2 Nitelikler

Şekil 5'te görülen 2. Küme %25.1 müşteriyi kapsamaktadır. Bu müşterilerin hepsi; cinsiyeti erkek olmakla birlikte, ağırlıklı olarak kampanyasız ürün alan, yaş ortalaması 36,29 ve kazandırdıkları ciro ortalaması 268,39'dur sonuçlar az da olsa veri seti ortalamasının üstünde ciro getirdiğini göstermektedir.

Şekil 6'da 3. Kümede müşteri oranı %12'dir. Bu müşterilerin hepsi; cinsiyeti kadın olmakla birlikte, ağırlıklı olarak kampanyasız ürün alan, yaş ortalaması 37,35'tir ve veri seti yaş ortalamasının üzerindedir. Ciro ortalaması ise 250,77'dir ve az da olsa veri seti ortalamasının altında ciro getirmiştir.



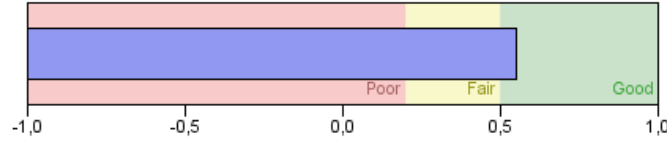
Şekil 6. Küme 3 Nitelikler



Şekil 7. Küme 4 Nitelikler

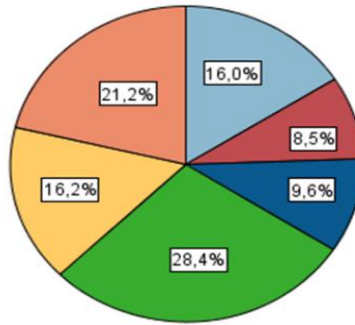
4. Kümede müşterilerin %21,7'si bulunmaktadır. Bu müşterilerin tamamının cinsiyetinin kadın olmasının yanında, ağırlıklı olarak kampanyalı ürün almaktadır, müşterilerin yaş ortalaması 37,73'tür ve 4. kümedeki müşterilerin yaş ortalaması veri seti yaş ortalamasının üzerindedir. Ciro ise ortalaması 250,5'dir.

İki Aşamalı kümeleme algoritmasında en iyi ikinci sonuç veren diğer örnekte nitelik seçimi; eğitim durumu, cinsiyet, ciro ve yaş olarak belirlenmiştir. Seçilen bu nitelikler sonucu Silhouette katsayısı 0,5 bulunmuştur.



Şekil 8. Örnek 2 Silhouette Katsayısına Endeksli Kümeleme Kalitesi

2. örnekte oluşan küme sayısı 6 olmuştur. Kümelerin dağılım yüzdeleri şekil 9'daki gibidir.

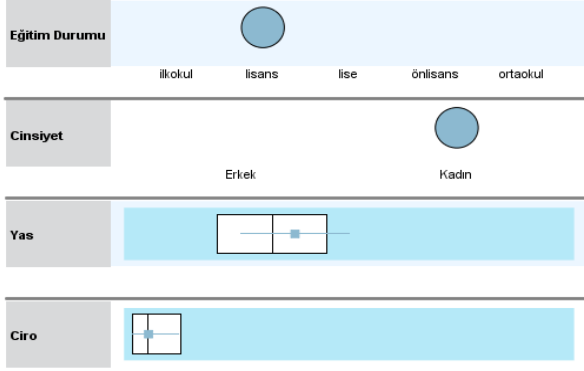


Şekil 9. Örnek 2 Kümelerin Dağılım Yüzdeleri

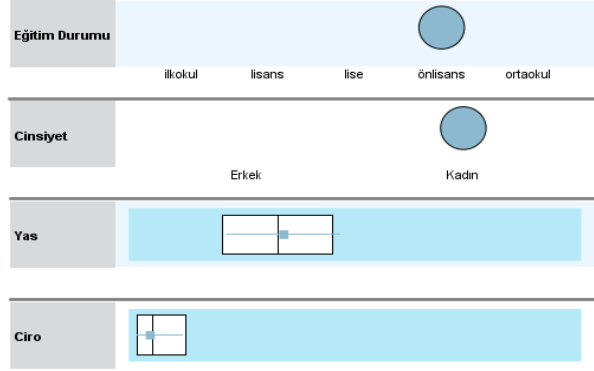
Tablo 3. Örnek 2 Küme içi Dağılımlar

Size	Inputs			
16,0% (2075)	Eğitim Durumu Lisans (100,0%)	Cinsiyet Kadın (100,0%)	Yaş 39,56	Ciro 236,35
8,5% (1108)	Eğitim Durumu önlisans (100,0%)	Cinsiyet Kadın (100,0%)	Yaş 36,81	Ciro 227,69
9,6% (1250)	Eğitim Durumu lise (87,5%)	Cinsiyet Kadın (93,8%)	Yaş 34,27	Ciro 211,71
28,4% (3689)	Eğitim Durumu lise (100,0%)	Cinsiyet Erkek (100,0%)	Yaş 34,81	Ciro 239,35
16,2% (2099)	Eğitim Durumu önlisans (100,0%)	Cinsiyet Erkek (100,0%)	Yaş 35,37	Ciro 230,58
21,2% (2747)	Eğitim Durumu lisans (100,0%)	Cinsiyet Erkek (100,0%)	Yaş 39,10	Ciro 240,93

Küme içi niteliklerin yüzdeler dağılımları Tablo 3'te gösterilmiştir. 1. Küme %16 oranında müşteriyi kapsamaktadır. Bu müşterilerin tümünün cinsiyeti kadın ve eğitim durumu lisans mezunudur, yaş ortalaması 39,56'dır ve ortalama yaşın oldukça üzerindedir. Ciro ortalaması ise 236,35'dir. Şekil 10'da 1. Kümeye ait nitelikler gösterilmiştir.

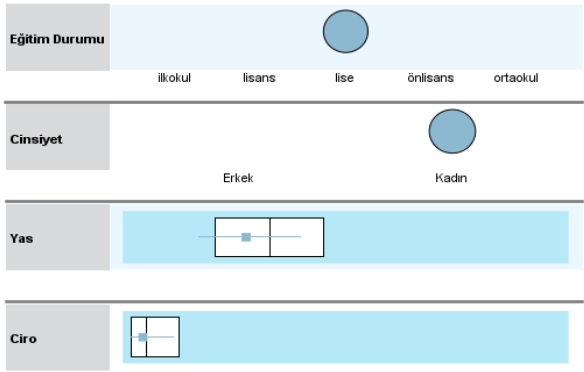


Şekil 10. Küme 1 Nitelikler

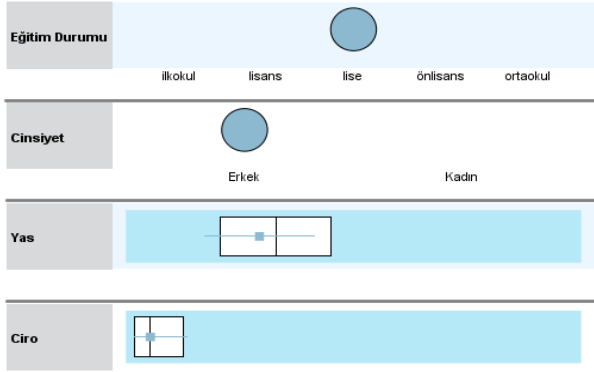


Şekil 11. Küme 2 Nitelikler

Şekil 11'de görülen 2. küme %8,5 müşteriye kapsamaktadır. Bu müşterilerin tümünün cinsiyeti kadın ve eğitim seviyeleri ön lisans düzeyindedir, yaş ortalamaları 34,27 ve veri seti ortalama yaşının üzerindedir. Elde edilen ciro ortalaması ise 227,69'dur ve az da olsa veri seti ortalamasının altında ciro getirisi vardır

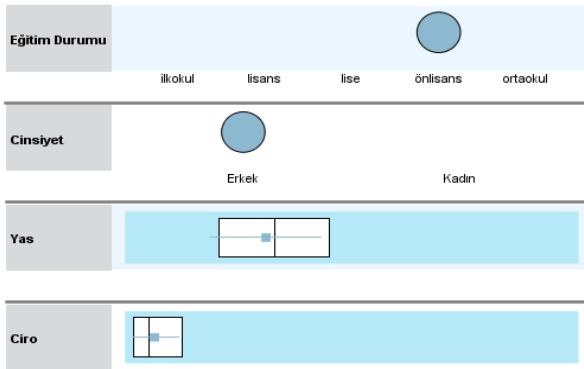


Şekil 12. Küme 3 Nitelikler

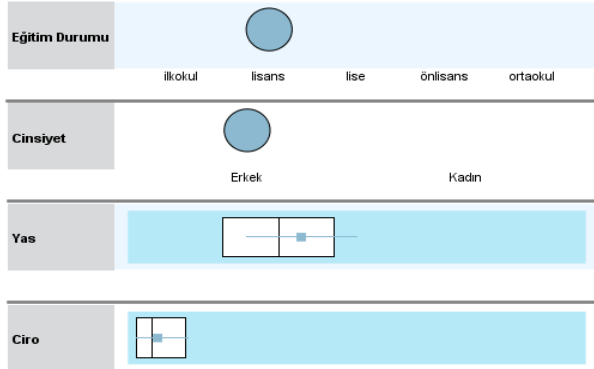


Şekil 13. Küme 4 Nitelikler

Şekil 12'deki 3. küme %9,6 müşteriye kapsamaktadır. Bu müşterilerin hepsinin cinsiyeti kadındır ve lise mezunlardır, yaş ortalaması 34,27'dir ve veri seti ortalamasının altındadır. Ciro ortalaması ise 221,71'dir ve az da olsa veri seti ortalamasının altında ciro getirileri mevcuttur. Şekil 13'de %28,4'lük müşteri oranıyla 4. küme yer almaktadır. 3. Kümedeki müşterilerin tümü lise mezunu erkeklerdir, yaş ortalamaları 34,81'dir ve veri seti ortalamasının altındadır. Ciro ortalaması ise 239,35'dir.



Şekil 14. Küme 5 Nitelikler



Şekil 15. Küme 6 Nitelikler

5. Küme şekil 14'te gösterilmektedir. Bu küme %16,2'lik müşteri oranına sahiptir. Erkek ön lisans mezunlarının oluşturduğu kümede, yaş ortalaması 35,37'dir ve veri seti ortalamasının altındadır. Ciro ortalaması ise 230,58'dir. 6. kümede müşteri oranı %21,2'dir. Bu müşteri ortalama 39,1 yaşında lisans mezunu erkeklerdir firmanın bu müşterilerden elde ettiği ciro ortalaması ise 240,93'dür ve veri seti ortalamasının üstündedir. 6. Kümeye ait grafiksel gösterim şekil 15'tedir.

3.2. Beklenti Maksimizasyonu Algoritması

Beklenti Maksimizasyonu (Expectation Maximisation) algoritması Witten ve diğ. (2016) tarafından geliştirilen Weka 3.8 : Machine Learning Software yazılımı kullanılarak uygulanmıştır. İki Aşamalı algoritmada en iyi silhouette katsayısı değerini veren 1. Örnek esas alınarak küme sayısı 4 olarak belirlenmiştir. 4 kümeli beklenti maksimizasyonu algoritması sonuçları tablo 4'te verilmiştir.

Tablo 4. Beklenti Maksimizasyonu Algoritması Kümeleri

Küme Yüzdeleri	0,18	0,15	0,41	0,25
Nitelikler	Küme 1	Küme 2	Küme 3	Küme 4
Yaş Ort.	38,11	37,37	35,81	37,10
Erkek	1.695,04	1.390,79	3.748,15	2.448,02
Kadın	823,00	758,97	2.050,52	1.080,51
Fiş Sayısı 1	583,02	767,44	5.797,67	3.526,87
Fiş Sayısı 2	880,02	1.382,32	1,00	2,00
Fiş Sayısı 3	735,00	1,00	1,00	1,00
Fiş Sayısı 4	234,00	1,00	1,00	1,00
Fiş Sayısı 5	66,00	1,00	1,00	1,00
Fiş Sayısı 6	15,00	1,00	1,00	1,00
Fiş Sayısı 7	7,00	1,00	1,00	1,00
Fiş Sayısı 8	4,00	1,00	1,00	1,00
Satış Adedi 1	1,00	27,99	5.767,74	2.425,25
Satış Adedi 2	1,00	2.121,77	30,93	1.103,28
Satış Adedi 3	1.369,00	1,00	1,00	1,00
Satış Adedi 4	675,00	1,00	1,00	1,00
Satış Adedi 5	319,00	1,00	1,00	1,00
Satış Adedi 6	77,00	1,00	1,00	1,00
Satış Adedi 7	37,00	1,00	1,00	1,00
Satış Adedi 8	27,00	1,00	1,00	1,00
Satış Adedi 9	10,00	1,00	1,00	1,00
Satış Adedi 10	4,00	1,00	1,00	1,00
Satış Adedi 11	6,00	1,00	1,00	1,00
Satış Adedi 12	2,00	1,00	1,00	1,00
Ciro Ort.	550,48	321,82	129,96	233,55
Lisans	1.033,02	805,20	2.106,88	1.286,91
Lisansüstü	70,01	79,13	175,68	130,18
Lise	852,01	718,42	2.183,02	1.365,55
Ortaokul	11,00	15,32	36,67	22,01
ÖnLisans	541,00	517,29	1.261,13	708,58
İlkokul	15,00	18,41	39,29	19,30
Kampanyalı	1.565,02	1.363,46	3.627,13	2.244,39
Kampanyasız	953,02	786,30	2.171,54	1.284,14

1. Kümedeki müşteri oranı %18'dir. Bu müşterilerin yaş ortalaması 38,11 olmakla birlikte 2/3'ü erkek 1/3'ü kadındır. Fiş sayısına bakıldığında en fazla 1 ve 2 kez alışveriş yapmış müşterilerin fazla olduğu fakat diğer kümelerle göre bu küme içindeki müşterilerin daha sık alışveriş yapan müşteriler olduğu anlaşılmaktadır. Diğer kümelerin aksine bu müşterilerin çoğunluğu 3,4, 5 ve daha fazla satınalma eylemi gerçekleştirmişken 1 ve 2 satınalma yapan müşteri sayısı 1'dir. Ortalama ciro 550'dir ve kümeler arası en yüksek ortalama ciro bu kümede gözükmemektedir. Müşterilerin eğitim durumu sırası ile lisans, lise ve önlisans olarak sıralanmaktadır. Son olarak çoğunlukta ağırlıklı olarak kampanyalı ürün alan müşteri sayısı fazladır.

2. Küme müşterilerin %15'ini kapsamaktadır. Bu kümeye mensup müşterilerin yaş ortalaması 37,37 olmakla birlikte 2/3'ü erkek 1/3'ü kadındır. Küme 1'den farklı olarak fiş sayısına bakıldığında yoğunlukta 1 ve 2 kez alışveriş yapmış müşterilerin fazla olduğu, bu alışverişlerinde 2 adet satış yapıldığı gözükmemektedir. Ortalama ciro 321'dir ve 2. olarak en yüksek ciroyu getiren kümedir. Eğitim durumları sırası ile lisans, lise ve önlisans olan müşterilerin sayısı fazladır. Çoğunlukta ağırlıklı olarak kampanyalı ürün alan müşteri sayısı fazladır.

3. Küme %41 müşteriyi kapsamakta olup en büyük yüzdeye sahip kümedir. Bu kümedeki müşterilerin yaş ortalaması 35,81 olmakla birlikte erkek müşteriler kadın müşterilerden 2 kat fazladır. Bu kümedeki 7 müşteri haricindeki herkes 1 kere alışveriş yapmıştır ve diğer satışları göz ardı edebilecek kadar küme müşterileri tek bir ürün almıştır denilebilir. Ortalama ciro 129,6'dır ve en düşük cironun elde edildiği kümedir. Eğitim durumları sırası ile lisans, lise ve önlisanstır. Çoğunlukta ağırlıklı olarak kampanyalı ürün alan müşteri sayısı fazladır.

4. Kümede müşteri yoğunluğu %25'tir. Bu müşterilerin yaş ortalaması 37,10 olmakla birlikte erkek müşterilerin oranı kadın müşterilerin 2 katından fazladır. Bu kümede de 1 kez alışveriş yapan müşteri sayısı oldukça yüksektir ve müşterilerin genel

olarak 1 veya 2 adet ürün aldıkları tespit edilmiştir. Ortalama ciro 233,55'dir. Lise ve lisans mezunu müşterilerin sayısı fazladır. Çoğunlukta ağırlıklı olarak kampanyalı ürün alan müşteri sayısı fazladır.

Sonuç ve Değerlendirme

Bu çalışmada bir tekstil sektöründe hizmet veren perakende zincir mağazasının alışveriş işlemlerinde kullanılan müşteri kartlarıyla elde ettiği müşteri verileri analiz edilmiştir ve müşteriler profillerine göre segmente edilerek her küme için ayrı pazarlama stratejileri sunulmuştur.

İki aşamalı kümeleme algoritması en yüksek silhoutte katsayı değerini veren 4 kümelik örnek 1'e bakıldığında tüm kümelerdeki müşterilerin ortalama ciro getirileri ve yaş ortalamaları birbirlerine çok yakındır. Bu tekstil firması için yaşı 35 ve 40 arası değişen müşteri profili çok yüksek tutar alışveriş yapmayan, 1 veya 2 üründen fazla ürün satın almayan bireylerden oluşmaktadır. Küme 1 ve 2 ye bakıldığında yüzdesel olarak erkek müşterilerin daha fazla kampanyalı ürün aldıklarını söylenebilir. Aynı oranlar kadın müşteriler için küme 3 ve 4'te de geçerli iken kadın müşterilerin yaş ortalaması erkek müşterilere göre daha yüksektir. %62 oranında kampanyaları takip eden ve 35 yaş üstü kesime hitap eden kampanyalı ürünler satışa sunulabilir. Yüzdesel olarak çevrimiçi mağazada alışveriş oranının oldukça düşük olmasını göz önünde bulundurursak, 35 yaş üstü için pratik mobil uygulamalar geliştirilerek ve bu kümelere uygun ürünlerde müşterilerine uygun kampanyalar düzenleyerek internet üzerinden çevrimiçi satış artırılabilir.

6 kümeli iki aşamalı kümeleme analizinde ise müşterilerin eğitim durumları hakkında bilgi sahibi olunabilir. Müşteriler çoğunlukta erkek olmakla birlikte kadın müşterilere göre daha yüksek maddi tutarda alışveriş yaptığı ciro ortalamalarından anlaşılmaktadır. En yüksek müşteri sayısına sahip 4. küme lise mezunu erkek müşterilerden oluşmaktadır, bu müşterilerin yaşları ortalama yaşın altındadır. Yapılan alışveriş cirosu ortalamasının üzerinde olması sebebiyle bu müşterilerin beklentilerini karşılamak adına en çok tercih edilen ürünlerin yeni modelleri veya kampanyaları oluştuğunda müşterilerle iletişime geçerek bildirim yapılabilir. Kümeleme işleminin sonucuna bakıldığında müşteri profili genel olarak lise, lisans ve önlisans mezunu müşteriler olduğu için teknolojiyi ve yenilikleri sık sık takip eden bir kitle olabileceği yorumu yapılabilir, kadın müşterilerin yoğunluğunu arttırmak için özel gün kampanyaları yapılabilir.

Beklenti maksimizasyonu algoritmasında çıkan sonuçlara bakıldığında %18 müşteri sayısına sahip 1. kümenin daimi müşterileri ve en çok ciro getiren müşterileri olduğu söylenebilir. 1'den çok alışveriş yapmış ve 1'den çok ürün alınmıştır. Bu müşterilerin fazla yoğunlukta bir kısmı erkek olup kampanyalı ürünleri tercih etmişlerdir. Bu kümedeki müşteriler kampanyaları takip ederek, kampanya dahilinde birçok ürün almışlardır. Bu müşterileri kaybetmemek ve sadakatlerini arttırmak için direkt kişilere özel kampanyalar sunulması yararlı olacaktır.

Küme 2'ye bakıldığında 2 kez alışveriş yapan ve alışverişlerinde 2 adet ürün alan müşteri sayısının fazla olduğu gözükmektedir. Ciro ortalamasının üstündedir, bu durumda bu müşterilerin kampanyalı ürünleri daha fazla alması sonucu veri setinin son 2 kampanya dahilinde oluşturulmuş olması ihtimaline karşın bu müşteri profilinin kampanyaları takip eden ve kampanya dahilinde alışveriş yapmak, bu firmaya alışverişi sıklaştırmak isteyen bir profili olduğu yorumu çıkartılabilir.

Bu kümelerin tam aksine 3. küme %40 ile en fazla müşteri sayısına sahipken ortalama ciro getirisi en az olan kümedir. Yaş ortalaması genel ortalamasının altındadır. Bu müşterilerin mağazaya 1 kez geldiği ve 1 adet ürün aldığı çıkan sonuçlardan anlaşılmaktadır. Bu küme fiyatları pahalı bulunduğu için yüksek oranda kampanyalı ya da kampanyasız en uygun fiyatlı ürünü almak adına mağazaya 1 kez gelmiş müşterilerdir. %40'lık müşteri kümesini kaybetmemek ve onları mağazaya tekrar getirebilmek için bu müşterilere de özel kampanyalar düzenlenip firmaya bağlanmaları sağlanabilir.

Küme 4 ise küme 3'e çok benzer özellikler göstermekle birlikte yine oldukça yüksek müşteri sayısına sahiptir. Aslında bu iki kümenin tek farkı 4. Kümede 2 adet ürün almış müşterilerin yoğunluğudur. Bunun sonucunda bu kümedeki müşteriler ile 3. Kümedeki müşterilerin pazarlama stratejileri birleştirilebilir. Kaçmakta olan %60'lık bir müşteri profili için firma kendini tanıtmalı ve müşterilerin satın aldıkları ürün fiyat ve modellerine uygun ürün satışlarında müşteriler ile iletişime geçilmelidir.

Tüm sonuçlar analiz edildiğinde erkek müşterilerin kadın müşterilere göre 2 kat fazla olması kadın müşterilerin satınalma davranışlarının tanımlanamadığı ya da firmanın ürün çeşidinin kadın müşterilere hitap etmemesinden kaynaklı kadın müşterileri kazanamadığı sonucu çıkarılabilir. Alışveriş sayısı düşük ve satış adedi çok az olan müşterileri kaybetmemek için firmanın bu müşterileri kendi markasına alıştırmak adına kişiye özel kampanyalar, markayı hatırlatacak reklamlar sunması gerekmektedir. Bunların yanı sıra 35-40 yaş lise lisans ve önlisans mezunu müşteri profili mevcuttur, bu müşterilere sadakat aşılacak adına en çok kullanılan sosyal platformlarda da reklam kampanyaları yapılabilir. Son olarak bilişim teknolojisi ve sosyal ağların bu kadar yoğun kullanıldığı bir zamanda çevrimiçi alışverişin %10'un altında olması bu alanda yatırımın gerekli olduğu anlamına gelmektedir.

Kaynakça

- Bacher, J., Wenzig, K., & Vogler, M. (2004). SPSS TwoStep Cluster-a first evaluation.
- Bahari, T. F., & Elayidom, M. S. (2015). An efficient CRM-data mining framework for the prediction of customer behaviour. *Procedia computer science*, 46, 725-731.
- Biscarri, F., Monedero, I., García, A., Guerrero, J. I., & León, C. (2017). Electricity clustering framework for automatic classification of customer loads. *Expert Systems with Applications*, 86, 54-63.
- Bradley, P. S., Fayyad, U., & Reina, C. (1998). Scaling EM (expectation-maximization) clustering to large databases.
- Chiu, T., Fang, D., Chen, J., Wang, Y., & Jeris, C. (2001, August). A robust and scalable clustering algorithm for mixed type attributes in large database environment. In *Proceedings of the seventh ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 263-268). ACM.
- Chong, A. Y. L., Ch'ng, E., Liu, M. J., & Li, B. (2017). Predicting consumer product demands via Big Data: the roles of online promotional marketing and online reviews. *International Journal of Production Research*, 55(17), 5142-5156.
- Claveria, O., & Poluzzi, A. (2017). Positioning and clustering of the world's top tourist destinations by means of dimensionality reduction techniques for categorical data. *Journal of Destination Marketing & Management*, 6(1), 22-32.
- Erevelles, S., Fukawa, N., & Swayne, L. (2016). Big Data consumer analytics and the transformation of marketing. *Journal of Business Research*, 69(2), 897-904.
- Fuentes, I., Nápoles, G., Arco, L., & Vanhoof, K. (2018, September). Customer Segmentation Using Multiple Instance Clustering and Purchasing Behaviors. In *International Workshop on Artificial Intelligence and Pattern Recognition* (pp. 193-200). Springer, Cham.
- Gordini, N., & Veglio, V. (2017). Customers churn prediction and marketing retention strategies. An application of support vector machines based on the AUC parameter-selection technique in B2B e-commerce industry. *Industrial Marketing Management*, 62, 100-107.
- Hol, V., & Sokol, O. (2017). Clustering retail products based on customer behaviour. *Applied Soft Computing*, 60(C), 752-762.
- Kumar, V., Rajan, B., Venkatesan, R., & Lecinski, J. (2019). Understanding the Role of Artificial Intelligence in Personalized Engagement Marketing. *California Management Review*, 61(4), 135-155.
- Kotler, P., Burton, S., Deans, K., Brown, L., & Armstrong, G. (2015). *Marketing*. Pearson Higher Education AU.
- Shih, M. Y., Jheng, J. W., & Lai, L. F. (2010). A two-step method for clustering mixed categorical and numeric data. *Tamkang Journal of science and Engineering*, 13(1), 11-19.
- Ünlü, R., & Xanthopoulos, P. (2019a). A weighted framework for unsupervised ensemble learning based on internal quality measures. *Annals of Operations Research*, 276(1-2), 229-247.
- Ünlü, R., & Xanthopoulos, P. (2019b). Estimating the number of clusters in a dataset via consensus clustering. *Expert Systems with Applications*, 125, 33-39.
- Witten, I. H., Frank, E., Hall, M. A., & Pal, C. J. (2016). *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann.

Extended Abstract

Aim and Scope

In this study, retail textile sales data were analyzed within the scope of the textile sector, which is one of the locomotive sectors of our country. Customer-oriented marketing strategies in addition to promotional and campaign work for customer groups using Two-Step Clustering Analysis method and Expectation Maximization algorithm was conducted clustering analysis to improve. Intelligent marketing proposals and strategies for the company are presented in accordance with the results of the customer segmentation.

Methods

Since the data set contains both categorical and numerical data, the Two Step Method is used in cases where two types of data are found in the same set of data. Most clustering algorithms can work on either complete pure numerical data or full pure categorical data. The two-step clustering algorithm is illustrated by experimental evidence that more robust results can be achieved in clustering mixed data. (Shih vd. , 2010)

Apart from the two-step clustering algorithm, the "Expectation Maximization" clustering method also uses the suitable. Expectation Maximization algorithm can be seen as generalization of density prediction clustering problems with a mixed model framework. An effective method for calculating mixed model parameters. The ability to model from a statistical point of view is stronger than other iterative techniques. (Bradley vd. , 1998)

Findings

In this study, customer data obtained by a retail chain store serving a textile industry with customer cards used in shopping transactions were analyzed and customer profiles were analyzed. and presented marketing strategies for each cluster separately.

Conclusion

When all results are analyzed, male customers are 2 times more than female customers, the purchasing behavior of women customers cannot be identified, or the company's product type can be found to female customers can be drawn from the fact that women can not win customers due to their inability to address. In order not to lose customers whose number of purchases is low and the number of sales is very low, the company offers personalized campaigns to accustom these customers to their brand, advertising to remind the brand. is required. In addition to these, 35-40 years old high school undergraduate and associate degree customer profile is available, advertising campaigns can be made on the most used social platforms in order to instill loyalty to these customers. Finally, the fact that online shopping is below 10% at a time of use of information technology and social networks means investment is necessary in this area.