

E-TİCARET VERİLERİNİN MÜŞTERİ PROFİLİ AÇISINDAN DEĞERLENDİRİLMESİ

Öğr. Gör. Turgut ÖZSEVEN

Gaziosmanpaşa Üniversitesi, Turhal MYO, (turgutozseven@gmail.com)

Yrd. Doç. Dr. Taner ERSÖZ

Karabük Üniversitesi, İşletme Fakültesi, (tanerersoz@karabuk.edu.tr)

ÖZET

Ticaretin elektronik ortamda gerçekleştirildiği e-ticaret, işletmelerin giderlerini azaltırken, zaman ve mekândan bağımsız her an her yerde alışveriş yapılabilmesine olanak sağlamakta ve kârlılığını artırmaktadır. Günümüzde her sektörde e-ticaret yoğun olarak kullanılmakta ve alışveriş verileri işletmelerin veritabanlarında saklanmaktadır. Ancak bazı işletmeler için bunlar veri olarak kalırken, bazı işletmeler bu veriyi satış stratejilerini geliştirmek ve müşteri özelliklerini tespit etmek için anlamlı bilgiye dönüştürmektedir. Bu çalışmada, e-ticaret hizmeti veren bir işletmeye ait satış verileri istatistiksel yöntemlerle incelenmiş, alışverişe etki eden kullanıcı özellikler tespit edilmiş ve müşteriler belirli özelliklere göre kümelemeye tabi tutularak kullanıcılar ve alışveriş alışkanlıkları hakkında bilgiler elde edilmiştir.

Anahtar Kelimeler: E-ticaret, K-means, Ki-kare, Cramer's V.

EVALUATION OF CUSTOMER PROFILE OF E-COMMERCE DATA

ABSTRACT

E-commerce was carried out in the electronically of trade, businesses while reducing costs, allow for shopping anytime anywhere and increase profitability without the concept of time and space. Today, e-commerce widely used in every sector and shopping data is stored in databases of businesses. However, some businesses are transforming this data into meaningful information to develop sales strategies and identify the characteristics of customers while data them for some businesses. In this study, e-commerce service providers to a business's sales data examined by statistical methods, user specifications have been identified which shopping acting and users and about shopping habits information have been obtained customers subjected to clustering based on certain characteristics.

Keywords: E-commerce, K-means, Chi-Square, Cramer's V.

1. Giriş

Bilişim teknolojilerinin gelişmesi ile birlikte iletişimin sınırları kalkmış, gerek şirketler gerekse de kullanıcılar açısından zaman ve mekân kavramı bitmiştir. Bu gelişmeler, bilginin ve bilgi teknolojilerinin değerini artırmış, ekonomik ve sosyal yaşamı derinden etkilemiş ve yeni bir ticaret kavramı olan Elektronik Ticaret (e-ticaret) kavramını ortaya çıkartmıştır.

E-ticaret, ticaretin elektronik ortamda yapılmasını ifade eder ve internet kullanımının artmasıyla ortaya çıkmıştır. E-ticaret ile birlikte müşteri ve firma kavramları yeni bir boyut kazanmıştır. Artık firmaya ait fiziki bir işyeri bulunmayıp ürünlerin tanıtımı ve pazarlaması internet üzerinden gerçekleştirilmekte ve müşteri portföyü de sadece işyerinin bulunduğu bölge ile sınırlı kalmayıp dünyanın tamamını kapsamaktadır.

E-ticaret'te sadece mal alımı söz konusu olmayıp, iş yönetimi, hizmet alımı, elektronik ortamdaki bilgi alışverişleri de e-ticaret kapsamı içerisinde yer almaktadır. E-ticarete kullanılan teknolojiler ve uygulamalar benzer olmasına rağmen kullanıcı taraflarına göre, firmalar arasında (B2B-Business to Business), firma tüketici arasında (B2C-Business to Customer), firma devlet arasında (B2G-Business to Government), müşteri devlet arasında (C2G-Customer to Government) e-ticaret gerçekleştirilebilmektedir (Taşhyan, 2006:81).

E-ticaret sistemleri üzerinde aynı ürün birden fazla işletme tarafından satışa sunulduğu için işletmeler farklı satış stratejileri geliştirerek rekabet güçlerini artırmaktadır. Bu amaçla veri madenciliği, firmalara müşteri ve ürün profilleri ile ilgili geleceğe ışık tutmaktadır. Böylece, işletmeler veri madenciliği sonuçlarına göre karar mekanizmaları oluşturmakta ve rekabet gücünü artırmaktadır (Vahaplar & İnceoğlu, 2001).

Gerek e-ticaret gerekse de farklı yöntemlerle işletmelerde depolanan veriler, yakın bir geçmişe kadar işletmelerin ilgisini çekmemekte ve sadece veri olarak kalmaktaydı. Ancak veri madenciliğinin gelişimi ve rekabet koşullarının ağırlaşması, gereksiz olarak görülen bu verileri değerli kılmış ve karar destek sistemleri açısından bilgiye dönüştürmeye başlamıştır. Günümüzde, gerek akademik gerekse de ticari anlamda veriler çok değerli bir duruma gelmiş ve sadece bu işlerle uğraşan işletmeler ortaya çıkmıştır.

E-ticaretin gelişimi ile birlikte işletmelerde oluşan büyük verilerin işletme kârını artırmak amacıyla yararlı bilgilere dönüştürülmesi gerekmektedir. Bu amaçla, e-ticaret verileri üzerinde veri madenciliği uygulamaları yapılması gerekmektedir. Veri madenciliği, büyük miktarda veri içinden gelecekle ilgili tahmin yapmamızı sağlayacak bağıntı ve kuralların bilgisayar programları kullanarak aranmasıdır (Ian & Eibe, 2005:8; Ersöz, 2013). E-ticaret verileri üzerinde veri madenciliği uygulanması ile potansiyel müşterilerin tahmini, müşteri alışkanlıklarının tespiti, müşteri odaklı kampanyalar düzenleme gibi işletmenin müşterisini tanıma ve kârını artırıcı çeşitli çalışmalar yapılabilir. Elektronik ticaretin altyapısı ile tüketicilerin tercihleri, alışkanlıkları ve demografik özellikleri takip edilebilir ve bu bilgiler kullanılarak ürün veya hizmet satın alanla, satan arasında kişiye özel ticari ilişki kurulabilir (Dolanbay, 2000:34).

Veri madenciliğinde kullanılan birçok program vardır. Bunlar arasında en çok tercih edilen programların başında IBM SPSS Modeler programı gelmektedir. Clementine olarak da

bilinen program CRISP-DM modeli çerçevesinde tahminleme modellerini geliştirebilen bir veri madenciliği yazılımıdır. Bu veri madenciliği programlarının dışında en çok kullanılan veri madenciliği programları; Weka 3.6.5, Statistica Data Mining&Predictive Analytics, RapidMiner, Tanagra, Oracle, Microsoft SQL Server Data Mining Add-ins for Microsoft Office, R Data Mining programlarıdır (Ersöz, 2013).

2. Literatür Taraması

E-ticaret işletmeler açısından yeni bakış açısı ve pazarlama politikası oluşturmasını gerektirir. Daha iyi müşteri yönetimi, pazarlama için yeni stratejiler, ürünlerin belirlenmesi için daha verimli operasyonlar sunar. Bu bakış açısını oluşturacak en önemli etken veri madenciliği araçlarıdır (Yuantao, 2008).

Vahaplar ve İnceoğlu (2001) tarafından sunulan bildiride; veri madenciliği konusunun elektronik ticaretteki uygulamaları vurgulanmış, veri madenciliğinin elektronik ticarete kazandıracığı faydalar belirtilerek, veri madenciliği teknikleri ve yöntemleri incelenmiştir. Shen vd. (2004), e-ticaret sitelerinde kullanıcı davranışlarını tahmin etmek için bir yaklaşım sunmuştur. Kalikov (2006) tarafından, bir yayınevi firmasının internet sitesindeki veriler dikkate alınarak, veri madenciliği birliktelik kuralları tekniği ile sepet ve sipariş tabloları incelenmiştir. Hangi ürünlerin kategorisinin değiştirilmesi gerektiği, kullanıcıların meslek ve ilgi alanı dağılımları, müşteri ilgi alanlarına göre satış grafikleri ve kullanıcıların ödeme seçenekleri ile ilgili bir veri madenciliği uygulaması gerçekleştirilmiştir. Yuantao (2008), e-ticaret verileri üzerinde hedef müşteri davranışlarının analizi için bir çalışma yapmıştır. Huang & Zhou (2009), e-ticarette kümeleme tabanlı veri madenciliği teknikleri üzerine bir araştırma yapmış ve k-means algoritmasının küme verimliliğini artırarak öneri hızını artıran bir algoritma sunmuşlardır. Cheng & Xiong (2009), e-ticarette web veri madenciliği konusu üzerinde durmuş, nesnelerin sınıflandırılması, veri madenciliği ile ağ verilerinin analizi ve e-ticaretin geleceği ile ilgili yazılımların geliştirilmesine yönelik referans bilgiler sunmuştur. Feng vd. (2012), e-ticarette modifiye ettikleri ID3 algoritması ile veri seti nitelik seçimi standartlarını geliştiren müşteri grubu temelli bir uygulama yapmıştır. Yadav vd. (2012), web kullanım madenciliğini kullanarak e-ticaret üzerinde kullanıcı davranışlarını araştırmış ve kullanıcıları yaş, cinsiyet, dil, davranış, ziyaret süresi gibi bilgiler üzerinde 6 kümeye ayırmıştır. Savaş vd. (2012), Topaloğlu ve Yılmaz tarafından veri madenciliğinin günümüz disiplinleri arasında geldiği noktaya değinilmiş ve Türkiye’de veri madenciliği üzerine yapılan çalışmalar ve gerçekleştirilen uygulamalar incelenmiştir. Nettleton vd. (2013), tarafından yapılan çalışmada internet üzerinden kitap satış yapan bir siteye ait arama kayıtlarını incelenmiş ve kitaplar ile kullanıcı davranışları arasındaki ilişki çizelgesel olarak gösterilmiştir.

Bu çalışmada, IBM SPSS Modeler yazılımı ile bir e-ticaret sitesine ait verilerden işletme müşterileri hakkında çeşitli bilgilerin ve ilişkilerin ortaya çıkartılması amaçlanmıştır. Bu amaçla, farklı şehirlerde cinsiyete ve yaşa göre hangi ürünlerin daha çok talep edildiği ve kaç adet ürün satın alındığına ilişkin veriler, çok değişkenli istatistik yöntemlerinden k-ortalamlar (k-means) yöntemi ile analiz edilmiş ve müşteri profilleri ortaya çıkarılmıştır.

3. Materyal ve Metod

Çalışmada kullanılan veriler bir e-ticaret sitesine ait veriler olup 2013 yılının son 6 ayını kapsamaktadır. Analizde 42188 adet satış verisi, Türkiye'nin farklı şehirlerinde yaşayan kişilerin yaş ve cinsiyetine göre dağılmaktadır. Ürünlerin verileri sağlayan işletme uygun görmediği için ismi açıklanmamıştır. İşletmeye ait verilerden sadece üç adet ürün grubuna (Kameralar, Spor Giyim, Mutfak Robotu) ilişkin veriler kullanılmıştır.

Veriler IBM SPSS Statistics programından çekilmiş, tanımlayıcı ve karşılaştırmalı istatistikler ve ilişki analizleri ortaya çıkarılmıştır. K ortalamalar yöntemi için veri madenciliği yazılımlarından IBM SPSS Modeler (Clementine) programı kullanılmıştır.

Çalışmada kullanılan değişkenlere ait frekans dağılımları, tanımlayıcı istatistikler, ki-kare ilişki analizleri, Cramer's v analizi, çoklu uyum analizi yapılmıştır.

Frekans dağılımı, verilerin düzenlenmesi, özetlenmesi ve anlaşılır hale getirilmesi için verileri gruplara ayırarak gözlenme sıklıklarını gösteren istatistiksel bir yöntemdir.

İlişki analizleri, kullanılan değişkenler arasındaki ilişkilerin belirlenmesinde kullanılır. Ki-kare analizi, frekans dağılımları üzerinden işlem yaparak değişkenlerin bağımsızlığını ölçmede kullanılır.

Cramer's v isimsel iki değişken arasındaki ilişkinin gücü hakkında bilgi verir. Bu katsayı iki değişken arasında ilişki olmadığı zaman 0'a eşit olur. Aksi durumda, Cramer v katsayısı 1'e eşit olsa da, değişkenler arasında yalnızca bir yönde mükemmel bir ilişki vardır (Bölükbaşı & Yıldıztan, 2009:345).

Çoklu uyum analizi, üç ya da daha fazla değişken arasındaki ilişki yapısını inceleyen, değişkenlerin yorumlanmasını kolaylaştıran, çapraz tablolarda satır ve sütunlarda yer alan değişkenler arasındaki ilişkileri grafiksel olarak gösteren bir yöntemdir (Suner & Çelikoğlu, 2010:43).

K-Means kümeleme, verilerin özelliklerine göre hiçbir küme bilgisi olmadan her verinin sadece bir kümeye ait olabilmesini sağlayan kümeleme yöntemidir (Davidson, 2002). Bu nedenle, keskin bir kümeleme algoritmasıdır. Küme içi benzerliğin yüksek, fakat kümeler arası benzerliğin düşük olması amaçlanır. Küme benzerliği bir kümedeki nesnelerin ortalama değeri ile ölçülmektedir, bu da kümenin ağırlık merkezidir (Han vd., 2001).

4. Bulgular

Ürün ve yaş değişkenlerine ait frekans dağılımları Tablo 1 'de verilmiştir.

Tablo 1: Yaşa Göre Ürün Dağılımı

		Ürün					
		Kameralar		Spor Giyim		Mutfak Robotu	
		N	%	N	%	N	%
Yaş	17 altı	136	21,6	413	65,6	81	12,9
	18-24	1232	19,3	4481	70,3	664	10,4
	25-34	4091	21,5	11388	60,0	3507	18,5
	35-44	2820	24,4	6590	57,0	2145	18,6
	45-54	971	26,8	2011	55,6	638	17,6
	55+	344	33,7	451	44,2	225	22,1
	Toplam	9594	22,7	25334	60,1	7260	17,2

Tablo 1'e göre alış veriş sitesi üzerinden incelenen üç kategori arasından tüm yaş aralıklarında en çok satın alınan ürün spor giyimdir. 55 ve üzerindeki müşterilerin diğer yaş aralıklarına göre alış veriş sitesi üzerinden daha az spor giyim satın aldıkları anlaşılmaktadır.

Cinsiyet ve ürün değişkenlerine ait frekans dağılımları Tablo 2'de verilmiştir.

Tablo 2: Cinsiyete Göre Ürün Dağılımı

		Ürün					
		Kameralar		Spor Giyim		Mutfak Robotu	
		N	%	N	%	N	%
Cinsiyet	Erkek	7981	24,0	19971	60,1	5279	15,9
	Bayan	1613	18,0	5363	59,9	1981	22,1
	Toplam	9594	22,7	25334	60,1	7260	17,2

Tablo 2'ye göre hem erkek hem de bayan müşterilerin her ikisinin de çok satın alınan ürün spor giyim olup erkeklerin ikinci olarak kameraları, bayanlarında mutfak robotu satın aldıkları anlaşılmaktadır.

Şehirlere ait frekans dağılım tablosunda çok fazla satır bulunduğu için gösterimde bölgeler kullanılmıştır. Bölgeler ve ürün değişkenlerine ait frekans dağılımları Tablo 3'te verilmiştir.

Tablo 3: Bölgelere Göre Ürün Dağılımı

		Ürün					
		Kameralar		Spor Giyim		Mutfak Robotu	
		N	%	N	%	N	%
Bölge	Karadeniz	1152	22,9	3096	61,5	785	15,6
	Akdeniz	1055	25,1	2521	60,0	623	14,8
	Marmara	3425	21,7	9473	60,1	2854	18,1
	İç Anadolu	1470	23,2	3724	58,7	1149	18,1
	Doğu Anadolu	583	21,7	1579	58,7	526	19,6
	Güneydoğu Anadolu	351	22,7	938	60,7	256	16,6
	Ege	1516	23,6	3887	60,6	1016	15,8
	Diğer	42	20,1	116	55,5	51	24,4
	Toplam	9594	22,7	25334	60,1	7260	17,2

Tablo 3'e göre tüm bölgelerde de spor giyim öncelikli olarak tercih edilmektedir. En çok ve en az satış yapılan 3 şehre ait ürün frekans dağılımı Tablo 4'te verilmiştir.

Tablo 4: En Çok ve En Az Satış Yapılan 3 Şehre Göre Ürün Dağılımı

		Ürün					
		Kameralar		Spor Giyim		Mutfak Robotu	
		N	%	N	%	N	%
En çok satış yapılan	İstanbul	1918	21,3	5185	57,6	1904	21,1
	Ankara	699	20,8	2023	60,2	640	19,0
	İzmir	614	20,9	1821	62,1	497	17,0
En az satış yapılan	Bayburt	13	22,0	37	62,7	9	15,3
	Siirt	11	14,5	52	68,4	13	17,1
	Ardahan	20	22,7	44	50,0	24	27,3

Tablo 4'e göre en çok satış yapılan şehir İstanbul ve en az satış yapılan şehir Bayburt'dur.

Alışveriş verilerinde yaşın ürün seçiminde herhangi bir etkisi olup olmadığını belirlemek amacı ile çapraz tablo ile karşılaştırma yapılmış ve veriler Tablo 5'te verilmiştir.

Tablo 5: Müşterilerin Ürün Tipleri ve Yaş Çapraz Tablosu

	Ürün-Yaş	17 altı	18-24	25-34	35-44	45-54	55+	Toplam
Kameralar	n	136	1232	4091	2820	971	344	9594
	Ürün İçindeki %	1,4	12,8	42,7	29,4	10,1	3,6	100,0
	Yaş İçindeki %	21,6	19,4	21,6	24,4	26,8	33,7	22,7
Spor Giyim	n	413	4481	11388	6590	2011	451	25334
	Ürün İçindeki %	1,6	17,7	45,0	26,0	8,0	1,8	100,0
	Yaş İçindeki %	65,5	70,2	60,0	57,0	55,6	44,2	60,1
Mutfak Robotu	n	81	664	3507	2145	638	225	7260
	Ürün İçindeki %	1,1	9,2	48,3	29,5	8,8	3,1	100,0
	Yaş İçindeki %	12,9	10,4	18,4	18,6	17,6	22,1	17,2
Toplam	n	630	6377	18986	11555	3620	1020	42188
	Ürün İçindeki %	1,5	15,1	45,0	27,4	8,6	2,4	100,0
	Yaş İçindeki %	100,0	100,0	100,0	100,0	100,0	100,0	100,0

Tablo 5'e göre müşterilerin % 45'i 25-34 arası yaş aralığında olup % 60'ı spor giyim ürünlerini satın aldığı ve müşterilerin % 27,4'ü 35-44 arası yaş aralığında olup, ürün tercihinde % 57'si spor giyim ürünlerini satın aldığı görülmektedir.

Yaş ile ürün değişkenleri arasında anlamlı bir ilişki olup olmadığını anlamak için Tablo 5'ten elde edilen verilere ki-kare testi uygulanmıştır. Sonuçlar Tablo 6'da verilmiştir.

Tablo 6: Yaş-Ürün Değişkenleri Ki-Kare Testi Sonuçları

	n	Serbestlik Derecesi	p (Önem düzeyi)
Anlamlılık	548,511	10	0,000
Olabilirlik Oranı	567,923	10	0,000
Doğrusal Bağlantı	3,495	1	0,062
Geçerli Veri Sayısı		42188	

Ki-kare tablosuna göre anlamlılık değeri $0,000 < 0,05$ olduğu için bireylerin yaş düzeyleri ile aldıkları ürün eğilim faktörü arasında anlamlı bir ilişkinin var olduğu görülmektedir.

Tablo 6'da verilen ki-kare testi sonuçlarına göre ilişkinin anlam düzeyini belirlemek için Cramer's v testi yapılmış ve sonuçlar Tablo 7'de verilmiştir.

Tablo 7: Yaş-Ürün Değişkenleri Cramer's v Testi Sonuçları

	Değer	P (Önem düzeyi)
Phi	0,114	0,000
Cramer's v	0,081	0,000
Veri Sayısı	42188	

Cramer's v testinden elde edilen değerlere göre; yaş ve ürün seçimi arasında ilişki olduğu, ancak ilişkinin derecesinin az olduğu (0,08) görülmektedir. Yaş arttıkça kamera satışlarının arttığı, spor giyim ürünleri satışının azaldığı görülmektedir.

Alışveriş verilerinde cinsiyetin ürün seçiminde herhangi bir etkisi olup olmadığını belirlemek amacı ile çapraz tablo ile karşılaştırma yapılmış ve veriler Tablo 8'de verilmiştir.

Tablo 8: Müşterilerin Ürün Tipleri ve Cinsiyet Çapraz Tablosu

Ürün-Cinsiyet		Erkek	Bayan	Toplam
Kameralar	n	7981	1613	9594
	Ürün İçindeki %	83,2	16,8	100,0
	Cinsiyet İçindeki %	24,0	18,0	22,8
Spor Giyim	n	19971	5363	25334
	Ürün İçindeki %	78,8	21,2	100,0
	Cinsiyet İçindeki %	60,1	59,9	60,0
Mutfak Robotu	n	5279	1981	7260
	Ürün İçindeki %	72,7	27,3	100,0
	Cinsiyet İçindeki %	15,9	21,1	17,2
Toplam	n	33231	8957	42188
	Ürün İçindeki %	78,8	21,2	100,0
	Cinsiyet İçindeki %	100,0	100,0	100,0

Tablo 8'e göre müşterilerin % 78,8'i erkek olup bunların % 60,1'i spor giyim ürünü satın alırken, müşterilerin % 21,2'si bayan olup bunların % 59,9'u spor giyim ürünü satın aldığı görülmektedir.

Cinsiyet ile ürün arasında anlamlı bir ilişki olup olmadığını anlamak için Tablo 8'den elde edilen verilere ki-kare testi uygulanmıştır. Sonuçlar Tablo 9'da verilmiştir.

Tablo 9: Cinsiyet-Ürün Ki-Kare Testi Sonuçları

	n	Serbestlik Derecesi	p (Önem düzeyi)
Anlamlılık	271,242	2	0,000
Olabilirlik Oranı	268,094	2	0,000
Doğrusal Bağlantı	266,599	1	0,000
Geçerli Veri Sayısı		42188	

Ki-kare tablosuna göre anlamlılık değeri $0,000 < 0,05$ olduğu için cinsiyet ile aldıkları ürün arasında anlamlı bir ilişkinin var olduğu görülmektedir.

Tablo 9’da verilen ki-kare testi sonuçlarına göre ilişkinin anlam düzeyini belirlemek için Cramer’s v testi yapılmış ve sonuçlar Tablo 10’da verilmiştir.

Tablo 10: Cinsiyet-Ürün Değişkenleri Cramer’s v Testi Sonuçları

	Değer	p(Önem düzeyi)
Phi	0,080	0,000
Cramer’s v	0,080	0,000
Veri Sayısı		42188

Cramer’s v testinden elde edilen değerlere göre; cinsiyet ve ürün arasında ilişki olduğu, ancak ilişkinin derecesinin az olduğu (0,08) görülmektedir. Yaş arttıkça kamera satışlarının arttığı, spor giyim ürünlerinin satışının azaldığı görülmektedir. Erkek müşteriler bayanlara göre daha fazla kamera satın alırken, bayan müşterilerin erkeklere göre mutfak robotunu daha fazla satın aldığı görülmektedir.

Alışveriş verilerinde bölgenin ürün seçiminde herhangi bir etkisi olup olmadığını belirlemek amacı ile çapraz tablo ile karşılaştırma yapılmış ve veriler Tablo 11’de verilmiştir. 7 bölge dışında kalan diğer veriler az sayıda olduğu için çapraz tabloda gösterilmemiştir.

Tablo 11: Müşterilerin Ürün Tipleri ve Yaşadığı Bölge Çapraz Tablosu

Ürün-Cinsiyet	A	B	C	D	E	F	G	Toplam	
Kameralar	n	1152	1055	3425	1470	583	351	1516	9552
	%*	12,0	11,0	35,9	15,4	6,1	3,7	15,9	100,0
	%**	22,9	25,1	21,7	23,2	21,7	22,7	23,6	22,8
Spor Giyim	n	3096	2521	9473	3724	1579	938	3887	25218
	%*	12,3	10,0	37,6	14,8	6,2	3,7	15,4	100,0
	%**	61,5	60,0	60,1	58,7	58,7	60,7	60,6	60,0

Tablo 11 devam

Mutfak Robotu	n	785	623	2854	1149	526	256	1016	7209
	%*	10,9	8,8	39,5	15,9	7,3	3,5	14,1	100,0
	%**	15,6	14,9	18,2	18,1	19,6	16,6	15,8	17,2
Toplam	n	5033	4199	15752	6343	2688	1545	6419	41979
	%*	12,0	10,0	37,5	15,1	6,4	3,7	15,3	100,0
	%**	100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0

* Ürün İçerisinde, ** Bölge İçerisinde

A: Karadeniz Bölgesi, **B:** Akdeniz Bölgesi, **C:** Marmara Bölgesi, **D:** İç Anadolu Bölgesi, **E:** Doğu Anadolu Böl.

F: Güneydoğu Anadolu Böl. **G:** Ege Bölgesi

Aalışveriř verilerinde bölge ve ürün seçimi etkileřiminin sonuçlarına göre; müşterilerin % 37,5'i Marmara bölgesinden olup ürün tercihinde % 60,1'ı spor giyim % 21,7'si kameraları tercih etmektedir. Müşterilerin % 15,3'ü Ege bölgesinden olup ürün tercihinde % 60,6'sı spor giyim % 23,6'sı kameraları tercih etmektedir.

Bölge ile ürün deęişkenleri arasında anlamlı bir iliřki olup olmadıęını anlamak için Tablo 11'den elde edilen verilere ki-kare testi uygulanmıřtır. Sonuçlar Tablo 12'da verilmiřtir.

Tablo 12: Bölge-Ürün Deęişkenleri Ki-Kare Testi Sonuçları

	n	Serbestlik Derecesi	p(Önem düzeyi)
Anlamlılık	534,300	162	0,000
Olabilirlik Oranı	529,231	162	0,000
Doęrusal Baęlantı	0,314	1	0,575
Geçerli Veri Sayısı		41979	

Ki-kare tablosuna göre anlamlılık deęeri $0,000 < 0,05$ olduęu için bölge ile bireylerin aldıkları ürün eğilim faktörü arasında anlamlı bir iliřkinin var olduęu görülmektedir. Tablo 12'de verilen ki-kare testi sonuçlarına göre iliřkinin anlam düzeyini belirlemek için Cramer's v testi yapılmıř ve sonuçlar Tablo 13'te verilmiřtir.

Tablo 13: Bölge-Ürün Deęişkenleri Cramer's v Testi Sonuçları

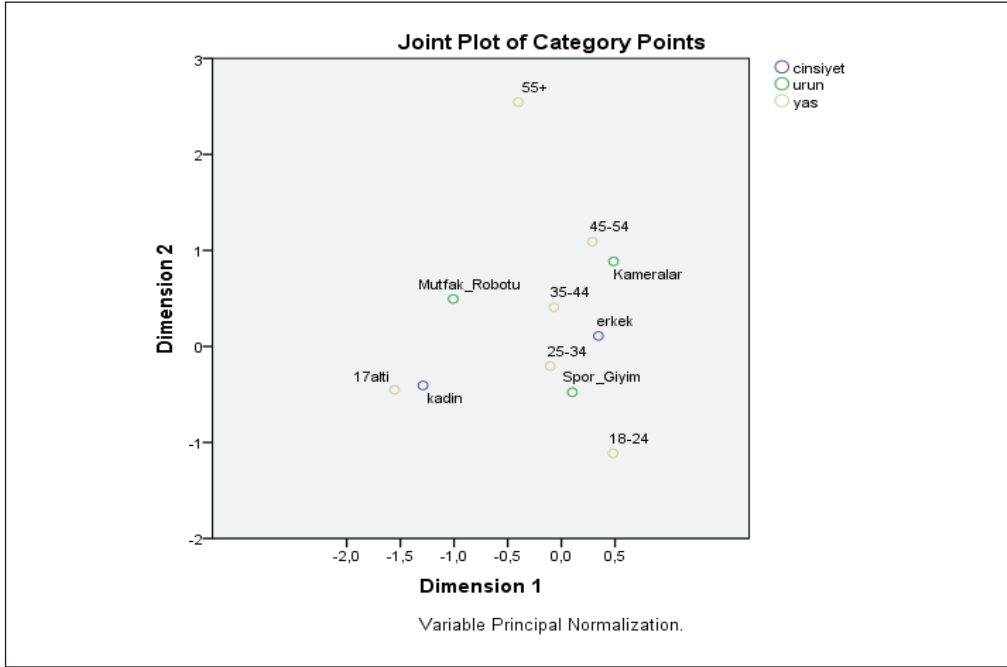
	Deęer	p (Önem düzeyi)
Phi	0,113	0,000
Cramer's V	0,080	0,000
Veri Sayısı		42188

Cramer's v testinden elde edilen deęerlere göre; bölge ve ürün seçimi arasında düşük seviyede iliřki olduęu (0,08) görülmektedir. Bölgelerin geliřmiřlik düzeyi ile e-ticaret üzerinden

yapılan alışveriş arasında doğrusal bir bağlantı olduğu görülmektedir. Tabi ki bunda bölgenin nüfus miktarı da ayrı bir etkidir. Ayrıca tüm bölgelerde spor giyim ürünlerinin satın alma oranı en yüksek olup, en düşük satın alma oranının mutfak robotunda olduğu görülmektedir.

Çalışmada kullanılan değişkenler çoklu uyum analizine tabi tutularak değişkenler arasındaki ilişki yapısı incelenmiş olup Şekil 1'de gösterilmiştir.

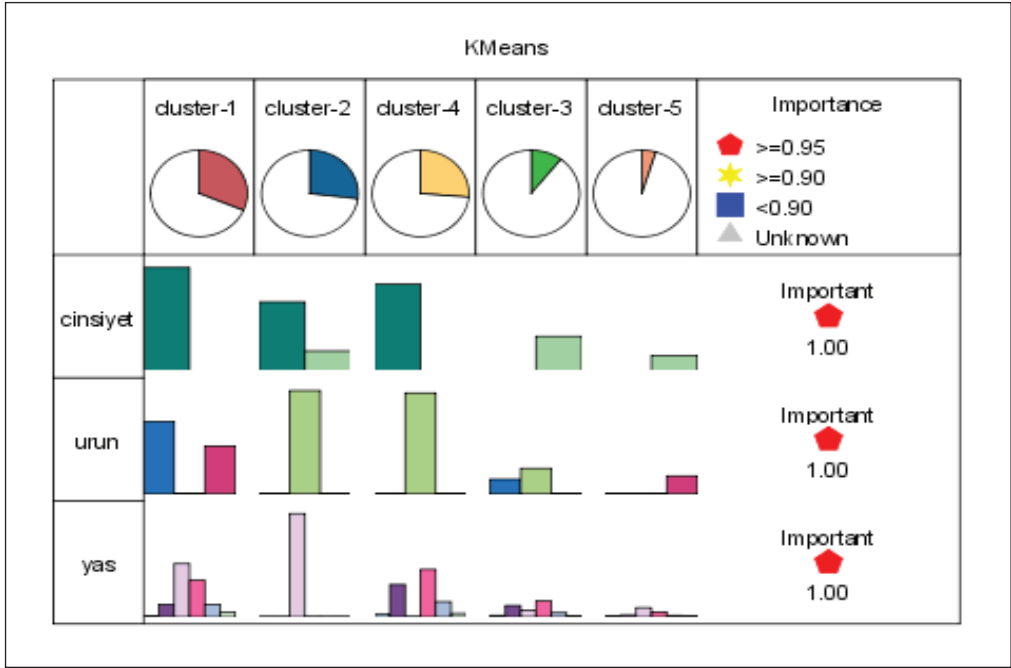
Şekil 1: Cinsiyet, Ürün ve Yaş Değişkenleri İçin Çoklu Uyum Analizi



Şekil 1 incelendiğinde; ürün alışverişinde 18-24 ve 25-34 yaş arasındaki erkek müşterilerin spor giyimi tercih ettikleri, 35-44 yaş arasındaki erkeklerin ise kameraları tercih ettikleri görülmüştür. Aynı şekilde de 17 yaş altı ve 18-24 yaş arasındaki kadınlarında spor giyimi tercih ettikleri, 35-44 yaş arasındaki kadınların ise mutfak robotu tercih ettikleri görülmüştür. Ürün tercihi öncelikli spor giyim ürünlerinde, sonrasında kameralarda yoğunlaşmıştır.

Çalışma verileri müşteri ve ürün özelliklerine k-means ile kümelemeye tabi tutulduğunda Şekil 2'de gösterildiği gibi 5 küme elde edilmiştir. Yaşanılan şehrin satın alınan ürün miktarında önemli olmadığı analiz sonucunda ortaya çıkmıştır.

Şekil 2. K-Means Kümeleme Analizi Sonuçları



Oluşan 5 kümeye ait özellikler Tablo 14’te verilmiştir.

Tablo 14: Elde Edilen 5 Kümeye Ait Özellikler

		1 Küme	2 Küme	3 Küme	4 Küme	5 Küme
Küme Kayıt Sayısı		13260	11388	4430	11129	1981
Cinsiyet	Erkek %	100	77,64	0	100	0
	Bayan %	0	22,36	100	0	100
Ürün	Kameralar %	60,19	0	36,41	0	0
	S. Giyim %	0	100	63,59	100	0
	M. Robotu %	39,81	0	0	0	100
Yaş	17 altı %	1	0	3,36	2,78	2,02
	18-24 %	10,32	0	28,31	31,91	10,3
	25-34 %	44,35	100	16,14	0	50,58
	35-44 %	30,42	0	39,5	47	27,26
	45-54 %	10,22	0	10,5	14,87	7,32
	55+ %	3,7	0	2,19	3,44	2,52

Elde edilen kümeleme sonuçlarına göre;

1. kümede, yaş aralığı 25-34 olan erkeklerin kamera satın aldıkları,
2. kümede, yaş aralığı 25-34 olan erkeklerin spor giyim ürünleri satın aldıkları,
3. kümede, yaş aralığı 35-44 olan bayanların spor giyim ürünleri satın aldıkları,
4. kümede, yaş aralığı 35-44 olan erkeklerin spor giyim ürünleri satın aldıkları,
5. kümede, yaş aralığı 25-34 olan bayanların mutfak robotu satın aldıkları görülmüştür.

K-means kümelemede her bir kümeyi oluştururken oluşan hatalar Tablo 15’te verilmiştir.

Tablo 15: K-Means Kümeleme Analizinde Oluşan İterasyon ve Hatalar

İterasyon	Hata
1	0,846
2	0,571
3	0,248
4	0,237
5	0,000

Tablo 15’e göre kümeleme 5 iterasyonda gerçekleşmiş ve 5.iterasyonda kümeler % 100 doğrulukla oluşturulmuştur.

5. Sonuç

Bilgi çağının başlaması ve bilişim teknolojilerinin gelişimi ile birlikte insanların internet kullanımını sadece ev ve ofis kullanımını aşmış, her zaman ve her yerde kullanılabilir hale gelmiştir. Bununla birlikte insanların alışveriş alışkanlıklarında da değişiklikler olmuş ve alışveriş sadece kurulu bir işletme üzerinde yapılmaktan daha geniş bir alana yayılmıştır. Bu gelişmeler sonucunda ortaya çıkan e-ticaret müşteriler açısından kullanım ve ulaşım kolaylığı sunarken, işletmeler açısından da daha az maliyetle dünyanın tamamına açılma imkanı sunmuştur. E-ticaretin ilk kullanıma başlandığı yıllarda kullanan müşteri sayısı ve işletme sayısının az oluşu rekabet açısından işletme için avantaj olurken müşteri açısından sınırlı seçenek anlamına gelmekteydi. Ancak günümüzde e-ticaret ciddi bir müşteri potansiyeline ve portföye sahiptir. Bu durum işletmeler açısından yeni rekabet ortamları oluşturmakta ve e-ticareti en verimli kullanan işletmeler bir adım öne geçmektedir. E-ticaret üzerinden alışverişte yüz yüze bir müşteri-satıcı diyalogu olmadığı için işletmeler müşterilerini tanımak için geçmiş alışveriş verilerini kullanmakta ve bu verileri rekabet güçlerini artırmak için bilgiye dönüştürmektedir. Bu çalışmada bir e-ticaret sitesine ait veriler 3 ürün grubu ve müşterilerin yaş, şehir, cinsiyet bilgilerine göre veri madenciliği ve istatistiksel analiz yöntemleri incelenmiştir. Elde edilen sonuçlara göre, müşterilerin büyük çoğunluğu spor giyim ürünlerini web sitesi üzerinden satın almayı tercih etmişken, en az tercih edilen ürün grubun mutfak robotu olduğu görülmüştür. Müşterilerin ürün tercihinde cinsiyet ve yaş bilgisi önem arz ederken, yaşadığı şehir bir anlam ifade etmemektedir. Şehir bilgisi ürün tercihi için anlamlı olmamasına karşın, müşteri potansiyeli açısından bilgi vermekte ve en çok satış yapılan bölge Marmara bölgesi olurken, en az satış yapılan bölge Güneydoğu Anadolu bölgesi olmuştur.

E-ticaret hizmeti veren işletmeler depoladıkları verileri bilgiye dönüştürmediği ve satış-müşteri profilini çıkartmadığı sürece, rakip işletmelerden her zaman bir adım geride kalacak ve sektördeki varlığını zamanla yitirecektir.

Kaynakça

- Böyükbaşı, A.G., & Yıldız, D.Ç. (2009). Yerel yönetimlerde iş tatminini etkileyen faktörlerin belirlenmesine yönelik alan araştırması. *Marmara Üniversitesi İ.İ.B.F. Dergisi*, 27(2), 345-366.
- Cheng, Y., & Xiong, Y. (2009, December). *Application of data mining technology in e-commerce*. 2009 International Forum on Computer Science-Technology and Applications, Washington, USA.
- Davidson, I. Y. (2002). *Understanding k-means non-hierarchical clustering*. Technical Report, Computer Science Department of State University of New York, Albany.
- Dolanbay, C. (2000). *E-ticaret strateji ve yöntemleri*. 1.Baskı, Ankara: Sistem Yayınları.
- Ersöz, F. (2013). *Veri madenciliği ve uygulamaları*. Ankara: Sage Yayıncılık.
- Feng Y., Huirnin Q., & Hemin J. (2012, August). *Study on the application of data mining for customer groups based on the modified ID3 algorithm in the e-commerce*. International Conference on Computer Science and Information Processing, Shaanxi, China.
- Han, J., Kamber, M., & Tung, A. K. H. (2001). *Geographic data mining and knowledge discovery*. USA: Taylor and Francis, Inc. Bristol.
- Huang W., & Zhou X. (2009, September). *Research of cluster-based data mining techniques in e-commerce*. Management and Service Science, Wuhan, China.
- Ian, W., & Eibe, F. (2005). *Data mining practical machine learning tools and techniques*. 2.Baskı, San Fransisco: Morgan Kaufmann Publishers.
- Kalikov, A. (2006). Veri madenciliği ve bir e-ticaret uygulaması. (Yüksek Lisans Tezi). Gazi Üniversitesi Fen Bilimleri Enstitüsü, Ankara.
- Nettleton, D. F., Baeza-Yates, R., & Marcos, M. C. (2013). Analysis of the user queries of an e-commerce bookstore in terms of the library of congress classification and key publishers. *Information Research-An International Electronic Journal*, 18(4).
- Savaş, S., Topaloğlu, N., & Yılmaz, M. (2012). Veri madenciliği ve Türkiye'deki uygulama örnekleri. *İstanbul Ticaret Üniversitesi Fen Bilimleri Dergisi*, 11(21), 1-23.
- Shen, L., Hawley, J., & Dickerson, K. (2004). E-commerce adoption for supply chain management in US apparel manufacturers. *Journal of Textile and Apparel, Technology and Management*, 4 (1), 1-11.
- Suner, A., & Çelikoğlu, C. C. (2010). Toplum tabanlı bir çalışmada çoklu uygunluk analizi ve kümeleme analizi ile sağlık kurumu seçimi. *Dokuz Eylül Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 25(2), 43-55.
- Taşlıyan, M. (2006). *Elektronik ticaret kavram ve uygulamaları*. Kahramanmaraş: Sakarya Kitabevi.
- Vahaplar, A., & İnceoğlu, M. (2001, Kasım). *Veri madenciliği ve elektronik ticaret*. Türkiye'de İnternet Konferansları, İstanbul.
- Yadav, M. P., Feeroz, M., & Yadav, V. K. (2012, July). *Mining the customer behavior using web usage mining in e-commerce*. Computing Communication & Networking Technologies, Allahabad, India.
- Yuanta J. S. Y. (2008, January). *Mining the e-commerce data to analyze the target customer behavior*. Knowledge Discovery and Data Mining, Las Vegas.