

Kargo Firmalarının Hizmet Kalitesinin Metin Madenciliği İle İncelenmesi^{1 2}

Batuhan ÇULLU³ - Algin OKURSOY⁴

Başvuru Tarihi: 15.11.2022

Kabul Tarihi: 25.04.2023

Makale Türü: Araştırma Makalesi

Öz

E-ticarette müşteri ile fiziksel temas sadece lojistik sürecin son aşaması olan kargonun teslimi adımıyla yaşanmaktadır. Bu nedenle kargo hizmet kalitesinde yaşanan sorunlar tüketici müşteri memnuniyetine yansımakta ve kötü deneyimler sosyal medya, e-ticaret siteleri ve tüketici blogları gibi çevrimiçi platformlara şikâyet olarak yansımaktadır. Bu çerçevede, tüketicilerin söz konusu platformda bıraktıkları negatif yorumların analiz edilmesi tedarik zinciri içerisindeki paydaşlar için kendi iş süreçlerinin geliştirilmesi ve rekabet güçlerinin artırılması adına önemli hale gelmiştir. Araştırmada Türkiye’de en yaygın hizmet ağına sahip dört kargo firmasına yapılan toplam 81.657 sayıdaki müşteri şikayeti www.sikayetvar.com adresinden Python programlama dili kullanılarak indirilmiştir. Bu yorumlar, metin veri analizi yöntemlerinden biri olan konu modelleme ile hizmet kalitesi boyutları altında sınıflandırılmıştır. Yorumların hizmet kalitesi boyutları altında sınıflandırılmasında, gözetimsiz sınıflama yöntemlerinden biri olan Gizli Dirichlet Ayrım (LDA) analizi kullanılmıştır. Çalışmanın amacı öncelikle, kargo sektöründe hizmet kalitesi boyutlarını oluşturan kelime gruplarının ortaya çıkartılması, sonrasında ise müşterilerin kargo hizmetinde sıkça sorun yaşadıkları hizmet boyutlarının açığa çıkartılmasıdır. Yapılan analizde 16 şikâyet konusu ortaya çıkartılmış ve en yüksek şikâyetin tüm firmalar için güvenilirlik hizmet kalitesi boyutundan geldiği saptanmıştır.

Anahtar Kelimeler: Metin Madenciliği, Hizmet Pazarlaması, Gizli Dirichlet Ayrımı

Atf: Çullu, B. ve Okursoy, A. (2023). Kargo firmalarının hizmet kalitesinin metin madenciliği ile incelenmesi. *Anadolu Üniversitesi Sosyal Bilimler Dergisi*, 23(2), 399-422.

¹ Bu çalışma etik kurul izin belgesi gerektirmemektedir.

² Bu çalışma Uluslararası Üretim ve Tedarik Zinciri Sempozyumunda (2022) sunulmuş ve sempozyum kitabında özet olarak basılmıştır.

³ Aydın Adnan Menderes Üniversitesi Söke İşletme Fakültesi Lojistik Yönetimi Bölümü, butuhan.cullu@adu.edu.tr, ORCID: 0000-0003-4969-1466

⁴ Aydın Adnan Menderes Üniversitesi Söke İşletme Fakültesi Lojistik Yönetimi Bölümü, a.okursoy@adu.edu.tr, ORCID: 0000-0003-2700-3503

Investigation of Cargo Companies' Service Quality Using Text Mining

Batuhan ÇULLU⁵ - Algin OKURSOY⁶

Submitted by: 15.11.2022

Accepted by: 25.04.2023

Article Type: Research Article

Abstract

Physical interaction with customers in e-commerce occurs only during cargo delivery, which is the final stage of the logistics process. As a result, poor cargo service quality is reflected in client satisfaction, and negative experiences are expressed as complaints on online platforms such as social media, e-commerce sites, and consumer blogs. In this context, monitoring negative comments published on the platform by consumers has become critical for supply chain stakeholders looking to enhance their own business processes and raise their competitiveness. A total of 81567 customer complaints filed to the four shipping businesses with the most extensive service network in Turkey were obtained using the Python programming language from the website www.sikayetvar.com. With subject modeling, one of the text data analysis methodologies, these remarks are categorized under the parameters of service quality. The unattended classification approach of Latent Dirichlet Allocation (LDA) analysis was utilized to classify comments into service quality characteristics. The purpose of this study is to first identify the word groups that make up the service quality dimensions in the cargo sector, and then to identify the service dimensions with which consumers frequently have issues. In total sixteen different complaint topics observed and collected under service quality dimensions.

Keywords: Text Mining, Service Marketing, Latent Dirichlet Allocation

⁵ Aydın Adnan Menderes University, Faculty of Business Administration, Logistics Management Department, butuhan.cullu@adu.edu.tr, ORCID: 0000-0003-4969-1466

⁶ Aydın Adnan Menderes University, Faculty of Business Administration, Logistics Management Department, a.okursoy@adu.edu.tr, ORCID: 0000-0003-2700-3503

Giriş

Son zamanlarda kargo pazarı dünya çapında dinamik bir şekilde gelişmektedir. Kargo hizmet sağlayıcıları müşterilere hızlı, takip edilebilir, taahhütlü, güvenilir ve esnek lojistik hizmetleri sunarak yalnızca firmalar için değil aynı zamanda bireysel müşteriler için de talep edilebilir bir sektör konumuna gelmişlerdir. Örneğin Türkiye’de 2015 – 2019 yılları arasında kargo pazarı %9 büyürken önümüzdeki 5 yılda %15 büyüme beklentisine sahiptir. Buna ek olarak 2020 yılında sektöre yapılan 10 milyar TL’lik yatırıma rağmen günlük kapasite eksikliği %30’dur. (“Türkiye Kargo Pazarı”, 2022). Aynı zamanda büyüyen pazar hacmine e-ticaretin artan popülaritesi de katkı sağlamaktadır.

Cronin ve Taylor'a (1992) göre, kargo hizmet kalitesinin iyileştirilmesi, rekabet avantajı elde etmek için bir zorunluluktur. Paket teslimat hizmetlerinde hizmet kalitesinin iyileştirilmemesi, hizmet sağlayıcıları için rekabetçi dezavantajlara yol açacaktır (Fabien, 2005). Bu nedenle, etkili hizmet kalitesine yönelik bir yaklaşım, kargo hizmet sağlayıcıları da dahil olmak üzere müşteri memnuniyetini kazanmak için tüm hizmet endüstrileri için bir önlemdir (Parasuraman vd., 1985). Firmaların sağladığı kargo hizmetinin kalitesi, müşteri tatmini, tüketicilerin tekrar satın alma davranışları (Büyükkelik vd. 2014), müşteri memnuniyeti (Songur, 2016) ve marka sadakati üzerinde de (Özgül vd., 2017) etkiye sahiptir. Bu nedenle yüksek rekabet koşulları altında kargo firmalarının yükselen pazar hacminden pay alabilmesi için müşteri tarafında algılanan hizmet kalitelerini sürekli olarak analiz etmeleri gerekmektedir.

Akademik yazında SERVQUAL modeli kargo veya parsiyel yük taşımacılığında hizmet kalitesinin ölçümü ve araştırılması için efektif bir araç olarak kullanılmaktadır. SERVQUAL, 1985 yılında Parasuraman ve diğerleri tarafından oluşturulan bir hizmet kalitesi modeli olup hizmet mükemmelliğinin değerlendirilmesi amacıyla kullanılan; fiziksel özellikler, güvenilirlik, heveslilik, güven ve empati boyutlarını içeren bir aygıttır (Parasuraman vd., 1985).

Aynı zamanda Parasuraman ve diğerleri (1988) ağızdan ağıza yayılan kullanıcı deneyimlerinin müşteri beklentilerini etkilediğini savunmaktadır. İnternetin gelişimi ve yayılması iletişim kanallarını artırmış ve insanların ürün ve hizmetler hakkındaki yorumlarını paylaşabilecekleri yeni çevrimiçi ortamlar sunmuştur. Günümüz müşterileri ürün ve hizmetler hakkındaki deneyimlerini şikayet günlükleri, sosyal medya, ürün siteleri, elektronik pazarlar, hızlı iletişim araçları vb. üzerinden çevrimiçi ortamlarda hızlıca paylaşmakta ve bu deneyimlerini daha fazla kişiye ulaştırabilmektedirler.

Müşteriler deneyimlerine ait video, ses, fotoğraf ve metinsel bilgileri de çevrimiçi ortamlarda paylaşmakta ve bu durum firmalar için değerli ve büyük bir veri kaynağı olmaktadır. Ancak çevrimiçi incelemelerin içeriği çoğunlukla anlaşılması ve yönetilmesi zor olan yapılandırılmamış veri tipinden oluşmaktadır (Qiao vd., 2017). Firmalar giderek artan müşteri verilerini inceleyerek hizmet kalitelerini, pazardaki durumlarını, müşteri taleplerini, müşteri segmentlerini, ürün yapılarını analiz edebilir ve daha doğru stratejilerle pazardan pay alabilirler. Bu nedenle, yapılandırılmamış verilerden ürün/ hizmet kusurları açısından değerli bilgilerin etkili ve verimli bir şekilde çıkarılması değer yaratımı için önem arz etmektedir (Qiao vd., 2017). Ancak yapılandırılmamış ve büyük miktarda gerçek zamanlı olarak güncellenen müşteri verisinin analizi için makine öğrenmesi, veri madenciliği, derin öğrenme gibi alanlara yönelmek gerekmektedir. Metin madenciliğinde kullanılan metotlardan biri olan Latent Dirichlet Allocation (Gizli Dirichlet Ayrım), doğal dil işlemede büyük bir belge kitaplığı içindeki gizli konuları ortaya çıkarmak için kullanılan istatistiksel bir modeldir. LDA genellikle belge sınıflandırma, konu modelleme gibi görevler için kullanılmaktadır

Çalışmada Türkiye’de faaliyet gösteren en yüksek hacme sahip ilk dört özel kargo firmasına müşteriler tarafından yapılmış 81.657 adet şikayet www.sikayetvar.com adresinden toplanmıştır. Elde edilen şikayetler metin formunda olup değerli kişisel bilgiler içermektedir. Çalışmanın amaçları; çok sayıda yapılandırılmamış müşteri yorum verisi üzerinde metin madenciliği yöntemlerinden Gizil Dirichlet Ayrımını (Latent Dirichlet Allocation – LDA) kullanarak;

- Kargo müşterilerinin şikayet konularını keşfetmek,
- Hizmet kalitesi boyutları altında şikayet konularını değerlendirmektir.

Araştırmanın temel çıktısı firma ve sektör bazında iç görü ve değerlendirme imkânı sunması olacaktır. Ek olarak araştırmanın, ulaşılabilen pazarlama literatürü içerisinde en yüksek sayıda Türkçe müşteri yorumunu LDA ile inceleyen çalışma olarak akademik yazına katkı sağlaması amaçlanmıştır.

Analiz sonucunda on altı şikayet konusu ve şikayetleri oluşturan kelimeler ortaya koyulmuştur. Sonrasında şikayet konuları yazarlar tarafından incelenerek hizmet kalitesi boyutları altında derlenmişlerdir.

Literatür Taraması

1980’lerde ürün ve hizmet kalitesi fikri pazarlama literatürüne girmiş ve fiziksel ürünler ve hizmetler olmak üzere iki kategoriye ayrılmıştır. Fiziksel ürünlerin pazarlanmasında kalite, artan bir hassasiyetle tanımlanmış ve ölçülmüştür; ancak, hizmetlerin pazarlanması söz konusu olduğunda, kalitenin tanımlanması, ölçülmesi ve kontrolünde zorluklar ortaya çıkmıştır. Hizmetler nesnel değil performanslar olduğundan, tek tip kalitede bir standart oluşturmak ve uygulamak çok zordur. Ancak, hizmet kalitesinin daha yüksek müşteri memnuniyeti ve karlılık gibi firmalara sağlayacağı net avantajlar nedeniyle, son kırk yılda araştırmacılar ve yöneticiler hizmet kalitesi alanına ilgi göstermişlerdir. Grönroos (1984) işletmelerin rekabetçi olabilmesi için tüketicilerin kaliteyi nasıl algıladıklarını ve hizmet kalitesinin nasıl etkilendiğini anlamaları gerektiğini savunmuştur. Yazar hizmet kalitesinin üç unsurunu teknik kalite, fonksiyonel kalite ve imaj olarak tanımlamış ve müşteri beklentileri ile işletme performansı arasındaki boşluk analizine dayalı bir hizmet kalitesi modeli oluşturmuşlardır. Parasuraman ve diğerleri (1985) tarafından geliştirilen boşluk modeli ise, kalite boyutları boyunca müşteri beklentileri ve işletmenin performansı arasındaki dengesizliklerin hizmet kalitesinin bir fonksiyonu olduğunu ortaya koymuştur. 1988 yılında ise Parasuraman ve diğerleri (1988) SERVQUAL modelini geliştirerek müşterilerin hizmet algısını etkileyen beş temel boyutu “fiziksel özellikler, empati, güvenilirlik, heveslilik, güven” olarak belirlemişlerdir. Tablo 1’de boyutlar ve tanımlamaları sunulmuştur.

Tablo 1
Hizmet Kalitesi Boyutları

Boyutlar	Tanımlamalar
Fiziksel Görünüm	Kullanılan donanım, personel ve hizmet verilen yerin fiziki görünümü
Güvenilirlik	Taahhüt edilen hizmetin kusursuz, güvenilir biçimde yerine getirme becerisi
Heveslilik	Müşterilere karşı hevesli ve yardımsever olma, hizmeti yerine getirme ve sorunları çözme isteği
Güven	Çalışanların bilgili, nazik olması, güven duygusu vermesi
Empati	Firmanın müşterilere kişisel ilgi göstermesi ve duyarlılığı

*SERVQUAL: a multiple-item scale for measuring consumer perceptions of service quality.” Parasuraman vd., Parasuraman, A., Zeithaml, V. A., Berry, L. (1988), Journal of Retailing, 64(1), 12-40 makalesinden yola çıkılarak türkçeleştirilmiş ve tablolatırılmıştır.

SERVQUAL modeli ile pek çok farklı hizmet sektörü üzerinde hizmet kalitesi araştırması yapılmıştır. Kargo sektörü için de literatürde çalışmalara ulaşmak mümkündür. Karadeniz ve Balcı (2014) bir kargo firmasından yola çıkarak algılanan hizmet kalitesinin marka sadakatine etkisini ölçerken SERVQUAL boyutlarını kullanmışlardır. Onurlubaş ve Gümüş (2020) hizmet kalitesi boyutlarının müşteri memnuniyetine olan etkisini ölçmüşlerdir. Benzer şekilde Büyükkeklik ve diğerleri (2014), kargo hizmet sağlayıcılarında kalitenin tüketici davranışını ölçtükleri çalışmalarında SERVQUAL boyutlarını kullanmışlardır. Tablo 2’de ulaşılabilen kargo firmalarında hizmet kalitesi üzerine yapılmış çalışmalar derlenmiştir.

Tablo 2

Ulaşılabilen Kargo Firmalarında Hizmet Kalitesi Üzerine Yapılmış Çalışmalar

Referans	Veri Toplama Metodu (Varsa Kaynak)	Gözlem Sayısı	Ölçek	Araştırma Sorusu
Deniz ve Gökmerdan (2011)	Anket	400	-	Müşterilerin Kargo Hizmetine Yönelik Tutum ve Düşünceleri
Yee ve Daud (2011)	Anket	103	SERVQUAL	Kargo Hizmet Kalitesi – Müşteri Memnuniyeti
Büyükkeklik vd. (2014)	Anket	257	SERVPERF	Kargo Hizmet Kalitesi – Müşteri Tatmini – Tekrar Satın Alma Davranışı
Karadeniz ve Balcı (2014)	Anket	399	SERVQUAL	Kargo Hizmet Kalitesi – Marka Sadakati
Gürce ve Tosun (2017)	Çevrimiçi Müşteri Yorumu (www.sikayetvar.com)	200	SERVQUAL	Kargo Hizmet Firmaları Müşteri Şikayetleri İçerik Analizi
Razik vd. (2018)	Anket	246	SERVQUAL	Kargo Hizmet Kalitesi – Müşteri Memnuniyeti
Topal ve Şahin (2019)	Anket	252	SERVQUAL	Kargo Hizmet Kalitesi – Müşteri Memnuniyeti
Burucuoğlu ve Yazar (2020)	Çevrimiçi Müşteri Yorumu (www.sikayetvar.com)	300	-	Kargo Firması Müşteri Şikayetleri İçerik Analizi
Onurlubaş ve Gümüş (2020)	Anket	384	SERVQUAL	Kargo Hizmet Kalitesi – Müşteri Memnuniyeti
Yıldız (2020)	Yüzyüze Görüşme	80	-	Kargo Firmaları Hizmet Kalitesi İçerik Analizi
Garip (2021)	Yarı Yapılandırılmış Yüzyüze Görüşme	6	SERVPERF/ SERVQUAL	Kargo Hizmet Kalitesi – Tüketici Ekstra Rol Davranışı
Atılğan ve Yoğurtçu (2021)	Çevrimiçi Müşteri Yorumu (www.twitter.com)	1616	-	Kargo Müşterilerinin Yorumlarından Duygu Analizi

Sunulan çalışmaların çoğu veri toplama yöntemi olarak anket kullanmış olup, sınırlı sayıda da olsa literatürde farklı veri toplama yöntemleriyle ve analiz yöntemleri ile yapılan araştırmalara ulaşmak mümkündür. Örneğin, Gürce ve Tosun (2017) www.şikayetvar.com üzerinden topladıkları 200 adet müşteri şikayetinden yola çıkarak yaptıkları içerik analizi çalışmasında en çok şikayet edilen hizmet kalitesi boyutlarını belirlemişlerdir. Yıldız (2020) ise 80 müşteri ile yüz yüze görüşmelerle topladığı verilerle içerik analizi gerçekleştirmiş ve kargo tüketicilerinin en fazla yaşadığı sorunları ortaya koymuştur.

İnternetin geniş kitlelere ulaşmasıyla beraber sanal ortamda yer alan platformlar aracılığıyla (sosyal medya uygulamaları, alışveriş, video-resim paylaşım, haber, kişisel blog siteleri vb.) her dakika sayısız veri türetilmektedir. Araştırmacılar bu verilerden bilgi keşfedebilmek ve bunların altında yatan insan düşüncesini otomatik olarak ortaya çıkarabilmek için güçlü metin madenciliği, makine öğrenmesi, derin öğrenme gibi veri bilimi yöntemlerine ihtiyaç duymaktadır. Örneğin, Atılğan ve Yoğurtçu (2021)'nin bir kargo firmasına ait 1138 Twitter gönderisini incelediği çalışmasında makine öğrenmesi yöntemlerinden duygu analizini kullandığı ve negatif yorumlarda sık geçen kelimeleri sundukları çalışması bu duruma örnek oluşturmaktadır. Aynı zamanda, Blei ve diğerlerinin (2003) ortaya koyduğu "Gizli Dirichlet Ayrım" metodu (Latent Dirichlet Allocation - LDA) metinsel veri içeren çok sayıda dokümanın içindeki konuları keşfetmeye yarayan bir yöntem olarak karşımıza çıkmaktadır.

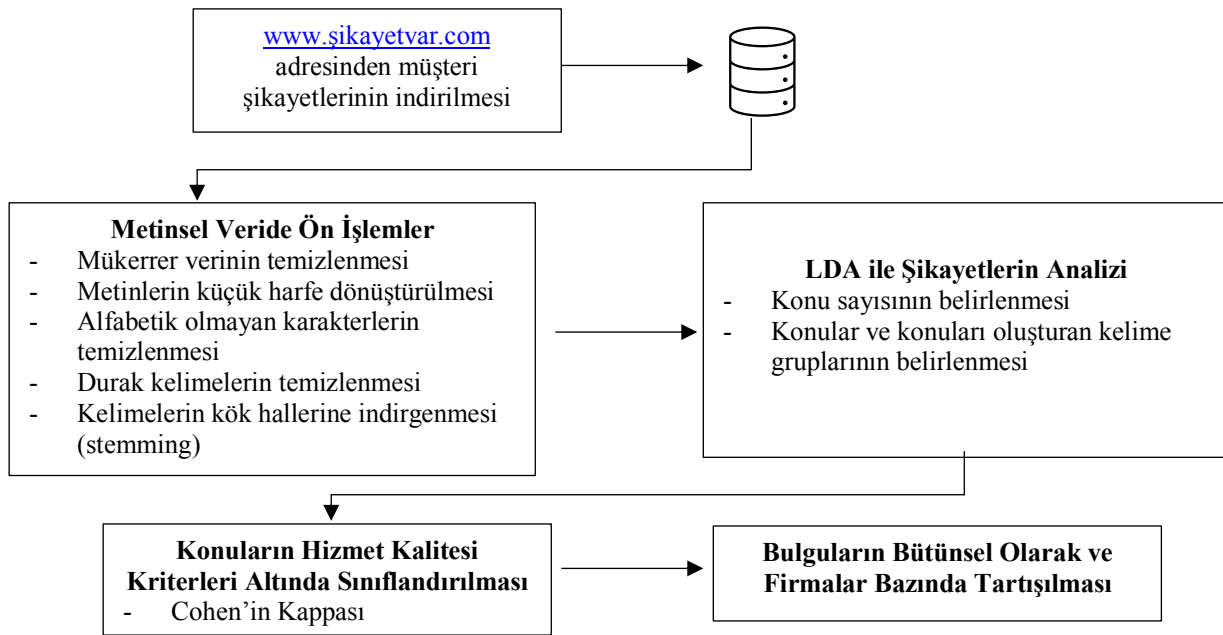
Sosyal bilimler alanında konu modellemesi ve özellikle LDA metodolojisi yabancı literatürde güncel bir metot olarak karşımıza çıkmaktadır. Sosyal bilimler alanında LDA metodunu kullanan öncül çalışmalardan birisi pazarlama alanında olup, Tirunillai ve Tellis (2014)'in bir firmanın çevrimiçi sohbet botundan aldıkları veriyle gerçekleştirdikleri stratejik marka analizidir. Wang ve diğerleri (2018) ise çevrimiçi müşteri yorumlarını LDA ile analiz ederek iki rakip ürünün pazar konumlandırma stratejilerini değerlendirmiştir. Poushneh ve Rajabi (2022) müşteri yorumları üzerinden yine müşterilerin beğenisini ölçtükleri çalışmasında LDA metodunu kullanmışlardır. Toubia ve diğerleri (2019) eğlence sektöründen film endüstrisindeki ürünlerin özelliklerini belirleyerek medya tüketim psikolojisini araştırmışlardır.

Bunlarla beraber müşteri yorumuna ulaşmanın görece daha kolay olduğu turizm alanında da LDA analizi araştırmacıların ilgisini çekmektedir. Örneğin, bu metot sayesinde çok sayıda tüketici yorumlarından müşterilerin otel (Guo vd., 2017, Sutherland vd., 2020), destinasyon (Putri ve Kusumaningrum, 2017), konaklama (Kiatkawsin vd., 2020) memnuniyetleri ve beklentileri araştırılmıştır.

Türkçe yazında LDA ile yapılan incelendiğinde sınırlı sayıda araştırmaya ulaşılmış ve çalışmaların çoğunluğunun mühendislik ve sağlık alanına ait olduğu görülmüştür. Sosyal bilimler alanında yapılan çalışmalarda ise yönetim bilişim sistemlerinin ağırlığı olduğu belirlenmiştir. Örneğin, Çallı ve diğerleri (2021) yönetim bilişim sistemleri disiplininde hazırlanan lisansüstü tezlerini, Kızılöz (2020) bilimsel makalelerin atıf sayısı tahminini, Aytuğ ve diğerleri (2020) üniversite bilgi sistemi servis destek taleplerini, Günay ve Güçdemir (2020) sivil toplum kuruluşlarının Twitter yorumlarını, Ala ve Uğuz (2021) girişimcilik, inovasyon ve Ar-Ge çalışmalarını analiz etmişlerdir. Türkçe literatürde LDA metodu kullanılarak gerçekleştirilmiş bir pazarlama çalışmasına ulaşılabilmiştir. Araştırmamız literatürde ulaşılabilen çalışmalar arasında en fazla sayıda veriyi metin madenciliği yöntemiyle analiz eden ilk çalışma özelliğini taşımaktadır. Analiz sonucunda ortaya koyulan şikayet başlıkları literatürde hizmet kalitesini ölçmekte sıklıkla kullanılan SERVQUAL boyutları altında incelenecektir.

Yöntem

Araştırma, kargo firmalarının hizmet kalitesine etki eden faktörlerini belirlemek amacıyla çevrimiçi müşteri şikayetlerini metin madenciliği başlıklarından konu modellemesinin altında yer alan gizli dirichlet ayrımı (LDA) ile analiz etmektedir. Bu doğrultuda oluşturulan araştırma süreci Şekil 1'de sunulmuştur.



Şekil 1. Araştırma Süreci

www.şikayetvar.com adresinden toplanan müşteri şikayetleri yoğun miktarda metinsel veri içermektedir. Araştırmanın amacına uygun bir şekilde metinsel verinin analiz edilebilmesi için konu modelleme yöntemlerinden LDA kullanılmıştır. LDA uygulamasından önce veri seti ön işlemlerden geçirilmiştir. Konu sayısının belirlenebilmesi için model tutarlılık değeri ve deneme yanılma yöntemleri kullanıldıktan sonra LDA analizi uygulanmıştır. Analiz sonucunda konular ve konuları oluşturan kelimeler belirlenmiştir. Ortaya çıkartılan konular, içerdikleri kelimeler ve şikayetler baz alınarak araştırmacılar tarafından hizmet kalitesi kriterleri altında sınıflandırılmışlardır. Araştırmacıların sınıflandırma tutarlılıkları tutarlılıkları Cohen'in Kappası ile ölçülmüştür. Ortaya çıkartılan konular, konulara ait kelimeler, konulara göre şikayet sayıları, hizmet kalitesine göre şikayet sayıları firma ve genel olarak bulgular bölümünde sunulmuştur. Sonuç bölümünde ise bulgular güncel literatür ile tartışılmış ve araştırmanın kısıtları, yönetsel ve akademik katkılar ile gelecek çalışmalar için öneriler sunulmuştur.

Konu Modelleme

Konu kavramı basitçe, bir kitap, paragraf, cümle, makale, haber vb. belgeler kümesinin ana fikri şeklinde tanımlanabilir. Buradan yola çıkarak konu, farklı metin seviyelerinde bireyler tarafından sezgisel olarak algılanan ve kolayca tanımlanamayan bir olgu olarak da tanımlanabilir. Metin madenciliği literatüründe konu, kelimeler ve belgeler arasında gizli ilişkileri bulmak için bir köprü olarak kullanılan; anlamsal ilişkiye sahip olma olasılıklı kelime kümeleridir (Daud vd., 2010). Bir belge çok sayıda kelimedenden oluşur ve kelimelerin olasılık dağılımlarından ibaret olan konular kolaylıkla modellenebilir, böylece tarama, arama ve belge benzerliğini değerlendirme gibi görevleri kolaylaştıran bir koleksiyon için yardımcı tanımlayıcı istatistiksel bilgiler elde edilebilir (Blei ve McAuliffe, 2007).

Literatürde, bireylerin gerek elektronik ortamda gerekse basılı kopya şeklinde oluşturduğu metinsel veriler için konu modelleme üzerine birçok yöntem önerilmiştir. Hemen hemen her alanda uygulanabilen bu yöntemlerle bireylerin gerçek dünyada yer alan konular üzerindeki düşünceleri öğrenilebilir ve bunlara göre zaman veya mekânsal değişikliklerde göz önüne alınarak pazarlama stratejileri, siyasi propaganda, tüketici davranış modelleri vb. geliştirilebilir. Konu modelleme sadece metin madenciliği alanında değil, değil, genetik bilimi, bilgisayar bilimleri ve sosyal ağlar gibi pek çok alanda başarıyla uygulanmaktadır (Lui vd., 2016).

Üretken bir model olan konu modelleme, belgelerin üretilebileceği basit bir olasılık sürecini belirtir (Steyvers ve Griffiths, 2007). Konu modellemenin ayırt edici özelliği, bir belge kitaplığının içeriğini konu adı verilen anlamlı kategorilere kodlayan otomatik bir süreci ifade etmesidir (Mohr ve Bogdanov, 2013). Metinsel veri içerisindeki noktalama işaretlerinin, kelimelerin anlamlarının ve konularının keşfedilen konulara etkisi olmamaktadır. Konu modelleme sürecinde, kelimelerin olasılık dağılımının konuları, söz konusu bu konulara ilişkin olasılık dağılımının da belgeleri oluşturduğu temel alınır (Barde ve Bainwad, 2017). Bir belgenin konu olarak temsili onun hakkında en önemli bilgileri içerir ve bu nedenle bilgi alma, kategorilere ayırma, özetleme ve metinleri bölme gibi birçok uygulama için değerlidir.

Buna göre bir belge yığını üzerinde konu analizi gerçekleştirmenin iki temel amacı vardır. Bunlardan birincisi belge kümesi içerisinde yer alan "k" konu sayısına karar vermek, diğeri de hangi belgelerin hangi konuyu ne ölçüde kapsadığını bulmaktır (Zhai ve Massung, 2016).

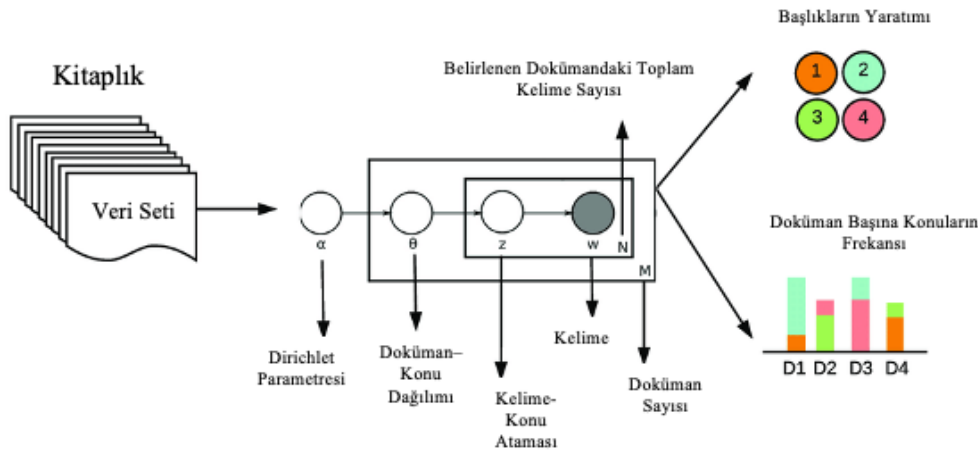
Konu modelleme için gizil semantik analiz (LSA), LDA, negatif olmayan matris faktörizasyonu (NMF), paralel gizli dirichlet ayrımı (PLDA), Panchinko tahsis modeli (PAM) gibi birçok makine öğrenmesi metodu kullanılabilir. Ancak akademik literatürde en sık kullanılan konu modelleme metodunun LDA olduğu görülmektedir (Eickhoff ve Neuss 2017).

Gizli dirichlet ayrımı (LDA)

Blei ve diğerlerinin (2003) ortaya koyduğu gizli dirichlet ayrımı, büyük bir belge kümesi içindeki konuları veya temaları keşfetmek için kullanılan bir tür istatistiksel konu modelleme yöntemidir.

LDA'nın arkasındaki temel fikir, her bir belgeyi (makale dahilinde ele alındığında müşteri yorumunu) konuların bir karışımı olarak ele almaktır; burada bir konu, kelimeler üzerinde bir dağılımdır. Örneğin, bir konu "spor" olabilir ve bu konu için kelimeler üzerindeki dağılım "futbol", "basketbol" ve "tarafar" gibi kelimeleri içerebilir.

LDA modeli her bir belgedeki sözcükleri üretmiş olma olasılığı en yüksek olan konuları ve her bir konuyla ilişkili olma olasılığı en yüksek olan sözcükleri öğrenmeye çalışır. Bunu, bir belgedeki her kelimenin konulardan biri tarafından oluşturulduğunu varsayarak yapar; konu seçimi belgedeki diğer kelimelerden ve tüm belgelerin toplandığı belge kitaplığındaki konuların genel dağılımından etkilenir.



Şekil 2. LDA Algoritmasının levha gösterimi

“Text Mining of Open-Ended Questions in Self-Assessment of University Teachers: An LDA Topic Modeling Approach” Buenano-Fernandez, D., Gonzalez, M., Gil, D., ve Luján-Mora, S., 2020., IEEE Access 8 makalesinden şekil alınmış ve yazarlar tarafından türkçeleştirilmiştir.

Şekil 2, algoritmanın işleyişini ayrıntılı olarak göstermektedir. LDA, bir dizi kelimedenden (N) oluşan her bir belgenin (M), gizli konular üzerinde Dirichlet'in olasılıksal bir dağılımı olarak temsil edilebileceğini varsayar. Burada α , konuya göre Dirichlet öncelik ağırlığını; Z, bir kelimenin belirli bir konuya atanmasını ve W, M belgesindeki gözlemlenen kelimeyi temsil eder.

Metinsel verileri ölçmek için, önce yorumlardaki konular belirlenmeli ve ardından yorumlar belirlenen konulara göre karakterize edilmelidir. Konu tanımlama için kullanılan LDA yaklaşımı, bir belgede kelime üretimini iki aşamalı bir süreç olarak ele alan olasılıklı konu modeli varsayımına dayanmaktadır (Blei, 2012; Blei, Ng ve Jordan, 2003, Jia, 2019):

- 1) Rastgele bir konu dağılımı seçilir.
- 2) Dokümandaki her bir kelime için,
 - a. (1)'inci formüldeki konuların dağılımından rastgele bir konu seçilir,
 - b. İlgili kelime dağılımından rastgele bir kelime seçilir.

Gerçek durumlarda ne konuların belgelere göre dağılımı ne de kelimelerin konulara göre dağılımı önceden bilinir; sadece dokümanlar gözlenir. Gizli ve gözlenen değişkenler arasındaki bağlantı, Denklem 1 kullanılarak ifade edilen ortak dağılımdır.

$$p(\beta_{1:K}, \theta_{1:D}, z_{1:D}, \omega_{1:D}) = \prod_{i=1}^K p(\beta_i) \cdot \prod_{d=1}^D p(\theta_d) \cdot \prod_{n=1}^N p(z_{d,n} | \theta_d) p(\omega_{d,n} | \beta_{1:K}, z_{d,n}) \quad (1)$$

Formülde K tüm konuları belirtirken, β_i , i 'inci konudaki kelimenin dağılımıdır. D belge kitaplığını gösterirken θ_d ise d dokümanındaki konuların oranlarıdır. N tüm kelimeleri sembolize ederken, z_d ise d dokümanındaki konu tahsisidir. $z_{d,n}$, d dokümanındaki n 'inci kelimenin konu tahsisini göstermektedir. Son olarak $\omega_{d,n}$ ise doküman d 'deki n 'inci kelimedir.

LDA'da konuların ve kelimelerin tanımlanması (Denklem 2) önceden bir sözlük oluşturulmadan, Gibbs örnekleme ile sonsal (posteriori) olarak yapılmaktadır.

$$p(\beta_{1:K}, \theta_{1:D}, z_{1:D} | \omega_{1:D}) = \frac{p(\beta_{1:K}, \theta_{1:D}, z_{1:D}, \omega_{1:D})}{p(\omega_{1:D})} \quad (2)$$

Basitleştirmek gerekirse LDA algoritmasının çalışması beş temel basamakta açıklanabilir. İlk olarak algoritma girdi olarak metin belgelerinden oluşan kitaplığı kullanır. İkinci adımda araştırmacı sabit bir konu sayısı belirler. Üçüncü basamakta kitaplıktaki her belge için, konuların olasılık dağılımı tahminlenir. Dördüncü adımda her konu için, kelimelerin olasılık dağılımı tahminlenir. Sonrasında Gibbs örneklemeyle, her belgedeki her kelime yinelemeli olarak gözden geçirilir ve belge için geçerli konu dağılımına ve konu için kelime dağılımına dayalı olarak ele alınan kelime için konu atanmasını günceller. Son olarak belirli sayıda

iterasyondan sonra algoritma, konuları yorumlamak ve yeni belgelere konu atamak için kullanılacak nihai konu-kelime ve belge-konu dağılımlarını çıkarır.

Modelin ürettiği sonuçlar Jia (2019)'un çalışmasında örneklendirilmiştir. Bu örneklendirme çalışmamıza uyarlandığında aşağıdaki gibi şekil almaktadır;

- Kargom ulaşmadı.
- Kargom iki haftadır gelmedi.
- Şube telefonlarına cevap vermiyor.
- Bilgi almak için şubeyle veya ilgili biriyle konuşamıyorum.
- Şubeyi aradığımda açmıyor ve günlerdir evde kargomu bekliyorum.

LDA bu cümlelere göre aşağıdaki sonuçları üretebilir;

- Birinci ve ikinci cümle: %100- 1. konu
- Üçüncü ve dördüncü cümle: %100- 2. konu
- Beşinci cümle: 67% – 1. konu; %33 – 2. Konu
- 1. Konunun kelimeleri: %30 kargo, %20 ulaş, %10 gel, %5 hafta... (konunun kargonun gecikmesi ile ilgili olduğu anlaşılabilir).
- 2. Konunun kelimeleri: %35 şube, %22 konuş, %10 ara, %5 aç ... (konunun iletişim problem olduğu anlaşılabilir).

Çalışmada LDA analizi için Python yazılım dili ve “Gensim” kütüphanesi kullanılmıştır (Rehurek ve Sojka, 2011). Kelimelerin köke indirgenmesi aşamasında ise “zemberek” kütüphanesinden yararlanılmıştır (Akın ve Akın, 2007).

Veri Seti

Türkiye’de en geniş pazar payına sahip dört kargo firmasından hizmet almış müşterilerin 2020-2022 yılları arasında www.şikayetvar.com adresine yazdıkları şikayetler ve şikayete ait bilgiler 13.01.2022 tarihinde yazarların Python yazılım dilinde oluşturdukları betik ile toplanmış ve MySQL veri tabanında saklanmıştır. Veri tabanı, müşteri şikayetlerini (metinsel veri), şikayet tarihlerini, kullanıcı isimlerini, şikayet bağlantısını içermekte olup; kullanıcıların kişisel bilgilerini ihtiva etmemektedir. Tablo 2’de firmalara göre toplanan şikayet sayıları sunulmuştur.

Tablo 3
Veri Seti Bilgileri

Firma	Yorum Sayısı	Veri Seti İçindeki Oranı
Firma A	21.605	%26
Firma M	21.663	%27
Firma S	19.106	%23
Firma Y	19.283	%24
Toplam	81.657	
Çalışmada Kullanılan	81.487	

Tablo 3’e göre dört kargo firması için yapılan 81.657 şikayet toplanmıştır. LDA kelime frekanslarından yola çıkarak olasılık değerleri oluşturduğu için başarılı sonuçlar alabilmek için metinsel veri ön işleme sokulmuştur. İlk olarak mükerrer olduğu saptanan 170 veri silinmiştir. Ardından tüm şikayetler küçük harflere

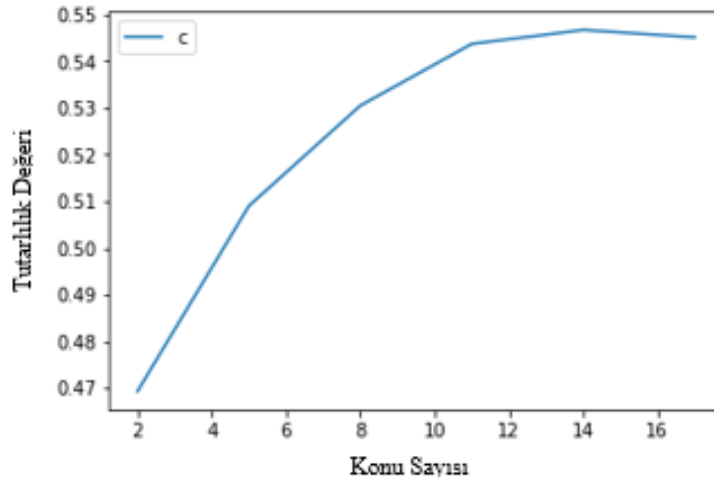
dönüştürülmüş, rakamlar ve noktalama işaretleri temizlenmiştir (örn: “1234 Numaralı Kargom!!” – “numaralı kargom”). Aynı zamanda günlük kullanımda sık geçen ancak anlam ifade etmeyen durak kelimeler (örn: ve, her, da vb.) ile firma adları şikayetlerden silinmiştir. Tüm kelimeler kök haline indirgendikten (örn: “gelmedi” – “gel”) sonra 81.487 adet ön işlemde geçirilmiş müşteri şikayeti belge kitaplığı olarak kabul edilmiştir.

Uygulama ve Bulgular

LDA algoritmasının uygulanabilmesi için konu sayısının belirlenmesi gerekmektedir. Ancak LDA için ideal konu sayısının belirlenmesi tartışmalı olup, akademik yazında birçok farklı metot önerilmiştir. Literatürde LDA kullanılarak yapılan konu modelleme analizlerinde konu sayısı genellikle deneme yanılma yöntemiyle belirlenmektedir (Blei vd., 2010; Blei vd., 2003; Kaplan ve Vakili, 2015, Bastani vd., 2019). Bu yöntemde araştırmacılar çeşitli konu sayılarını denemekte ve insan yargısına anlamlı kelime grupları oluşturan konu sayısını modelde kullanmaktadırlar.

Konu sayısı belirlemek için kullanılan bir diğer metot ise tutarlılık (Coherence) değeridir (Stevens vd., 2012). Bu çalışmada konu sayısını belirlemek için Röder ve diğerleri (2015) tarafından LDA algoritması için önerilen bir tutarlılık değeri (makalelerinde CV olarak anılmaktadır) ve araştırmacıların yargıları kullanılmıştır. CV değeri maksimum tutarlılığı temsil eden 1 (en yüksek) ile 0 (en düşük) arasında değişmektedir (Padarian vd., 2020).

Şekil 3'te görüldüğü üzere iki ile on sekiz konu aralığında çift sayılarda eğitilen LDA modelleri için CV değerleri hesaplanmış ve on dört konu sayısından sonra model tutarlılığının düşmeye başladığı gözlemlenmiştir. Yazarların da model çıktılarını değerlendirmesi ile müşteri şikayetlerinin on altı konu altında incelenmesine karar verilmiştir.



Şekil 3. LDA Modellerinin Konu Sayısına Göre Tutarlılık Değerleri

LDA bir gözetimsiz öğrenme algoritması olduğundan ortaya çıkarttığı konuları otomatik olarak etiketlendirmemektedir. Konuların tutarlılığı ve önemini değerlendirmek ve ardından bu değerlendirmelere dayanarak konulara etiket atamak için insan yargıları ve müdahalesi gerekmektedir (Chang vd., 2009). Bu nedenle her bir araştırmacı konuların içerisindeki kelimeleri ve konu altında geçen müşteri şikayetlerini inceleyerek konuyu temsil eden hizmet kalitesi boyutlarını belirlemiştir. Örneğin, dördüncü konu altında bulunan 'telefon', 'müşteri', 'şube', 'hizmet', 'cevap', 'ulaş', 'kapat', 'temsil', 'bekle', 'sistem' kelimeleri ve konunun baskın olduğu müşteri yorumları incelendiğinde şikayetlerin müşteri hizmetleri veya şube ile iletişim kuramama problemine dayandığı yargısına varılmıştır. Bu nedenle araştırmacı tarafından dördüncü konuya,

hizmet kalitesi boyutlarından 'heveslilik' atanmıştır (bkz: Tablo 3). Araştırmacıların kendi aralarındaki atama uyumları Cohen'in Kappası ile değerlendirilmiş ve ilk atamada 0.812 uyum sağlamıştır. Atamaların farklılık gösterdiği konular araştırmacılar tarafından tekrar değerlendirilmiş ve ortak yargıyla kabul edilmiştir.

Ek olarak, LDA'ya göre bir şikayet birden fazla konu içerebilmektedir. Örneğin, bir müşteri yazdığı şikayette öncelikle kargonun yanlış şubeye tesliminden bahsederken şikayetin devamında çözüm bulmak için aradığı şubeye ulaşamamasından yakınıyor olabilir. Ancak şikayette geçen kelime dağılımına göre hangi konunun şikayette baskın olduğu belirlenebilmektedir. Buna göre şikayet sayısı baskın konuya göre sayılarak belirlenmiştir. LDA analizi sonucunda ortaya çıkan on altı adet konu, konuların en yüksek benzerlik değerine (BD) sahip on kelimesi, konuya atanan hizmet kalitesi boyutu, şikayet sayısı ve belge kitaplığı içerisinde şikayet konusunun oranı Tablo 4'te sunulmuştur.

Tablo 4

LDA Çıktıları

Kelimeler	BD	Hizmet Kalitesi Boyutu	Şikayet Sayısı	Belge kitaplığı İçindeki Oran	Kelimeler	BD	Hizmet Kalitesi Boyutu	Şikayet Sayısı	Belge kitaplığı İçindeki Oran
KONU 1					KONU 2				
'sipariş'	0.545	Güvenilirlik	6029	%7.40	'iade'	0.195	Heveslilik	5661	%6.95
'firma'	0.096				'geri'	0.088			
'trendyol'	0.071				'para'	0.070			
'teslimat'	0.035				'gönder'	0.066			
'site'	0.026				'iste'	0.054			
'üzeri'	0.022				'ücret'	0.050			
'alışveriş'	0.021				'firma'	0.042			
'satıcı'	0.018				'satıcı'	0.021			
'taraf'	0.018				'yolla'	0.017			
'hepsiburada'	0.017				'ödeme'	0.015			
KONU 3					KONU 4				
'adres'	0.264	Güvenilirlik	5125	%6.29	'telefon'	0.185	Heveslilik	5325	%6.53
'şube'	0.121				'müşteri'	0.163			
'yanlış'	0.054				'şube'	0.117			
'gönder'	0.047				'hizmet'	0.116			
'alıcı'	0.047				'cevap'	0.053			
'gönderi'	0.036				'ulaş'	0.027			
'hata'	0.031				'kapat'	0.022			
'bulun'	0.027				'temsil'	0.021			
'doğru'	0.022				'bekle'	0.018			
'yeter'	0.018				'sistem'	0.015			

KONU 5				KONU 6					
'numara'	0.172	Güvenilirlik	4639	%5.69	'şikayet'	0.088	Heveslilik	4058	%4.98
'takip'	0.114				'talep'	0.077			
'bilgi'	0.103				'ilgi'	0.068			
'taraf'	0.093				'ilet'	0.047			
'gönderi'	0.086				'konu'	0.046			
'tarihi'	0.079				'taraf'	0.041			
'sistem'	0.019				'söyle'	0.040			
'öğren'	0.015				'dönüş'	0.039			
'ilgi'	0.015				'oluş'	0.033			
'ilet'	0.015				'bulun'	0.032			
KONU 7				KONU 8					
'çalış'	0.131	Empati	5124	%6.29	'paket'	0.063	Güvenilirlik	5983	%7.34
'söyle'	0.044				'gönder'	0.046			
'konuş'	0.037				'şekil'	0.040			
'personel'	0.029				'üzeri'	0.038			
'arkadaş'	0.027				'koli'	0.019			
'saygı'	0.024				'kutu'	0.018			
'şekil'	0.018				'taşı'	0.016			
'şikayet'	0.017				'hasar'	0.016			
'eleman'	0.015				'adet'	0.015			
'davran'	0.013				'zarar'	0.013			
KONU 9				KONU 10					
'firma'	0.122	Heveslilik	4307	%5.29	'şube'	0.221	Güvenilirlik	4487	%5.51
'sorun'	0.094				'getir'	0.103			
'aynı'	0.090				'teslimat'	0.054			
'yaşa'	0.072				'alın'	0.035			
'zaman'	0.067				'ilçe'	0.023			
'sürekli'	0.041				'uzak'	0.020			
'alışveriş'	0.037				'merkez'	0.020			
'durum'	0.026				'bura'	0.020			
'sıkıntı'	0.017				'bulun'	0.016			
'problem'	0.014				'bölge'	0.015			

KONU 11				KONU 12					
'konu'	0.036	Güvenilirlik	4254	%5.22	'şirket'	0.086	Empati	4569	%5.61
'işlem'	0.034				'insan'	0.055			
'ilgi'	0.032				'çalış'	0.044			
'durum'	0.031				'sorumsuz'	0.038			
'taraf'	0.024				'kötü'	0.035			
'kayıp'	0.024				'anla'	0.031			
'kaybol'	0.021				'gerçek'	0.029			
'yetki'	0.021				'mağdur'	0.027			
'fatura'	0.021				'yazık'	0.025			
'sonuç'	0.020				'yalan'	0.014			
KONU 13				KONU 14					
'bırak'	0.090	Güven	5148	%6.32	'ulaş'	0.195	Güvenilirlik	6207	%7.62
'kapı'	0.085				'elim'	0.058			
'kişi'	0.047				'sipariş'	0.045			
'site'	0.042				'lütfen'	0.034			
'getir'	0.039				'istanbul'	0.032			
'söyle'	0.033				'tarihi'	0.031			
'isim'	0.032				'gözük'	0.028			
'görev'	0.027				'yardım'	0.025			
'güven'	0.024				'aktar'	0.023			
'iste'	0.022				'nere'	0.022			
KONU 15				KONU 16					
'dağıtım'	0.211	Güvenilirlik	5446	%6.68	'teslim'	0.102	Güven	5125	%6.29
'mesaj'	0.177				'başka'	0.091			
'saat'	0.128				'kişi'	0.085			
'bugün'	0.061				'isim'	0.055			
'bekle'	0.042				'nere'	0.043			
'günü'	0.036				'göster'	0.034			
'çıkar'	0.034				'alan'	0.032			
'sabah'	0.023				'görün'	0.027			
'çıktı'	0.018				'mesaj'	0.025			
'tekrar'	0.018				'tanı'	0.023			

Sonuçlar detaylı olarak ele alındığında konu 14'ün tüm şikayetlerin %7.62'sine denk geldiği ve en sık gözlenen şikayet konusu olduğu ortaya çıkmaktadır. Konunun içerdiği kelimeler incelendiğinde ("ulaş", "elim", "sipariş", "lütfen", "nere" vb.) kargo ulaştırmada yaşanan problemler olarak görülmektedir. İkinci en yüksek şikayeti alan konunun %7.40 ile içinde çevrimiçi alışveriş sitelerinin geçtiği konu-1 olduğu ortaya çıkmıştır. Buradan çevrimiçi alışveriş teslimatında yaşanan problemlerin ayrı bir konu altında toplandığı ve en çok şikayet alan ikinci konu olduğu anlaşılmaktadır. Üçüncü sık gözlemlenen şikayetin ise hasarlı paket ve koli kelimelerini ihtiva eden konu 8'dir. Konu 3 ve konu 16'nın aynı sayıda şikayeti (5125 - %6.29) altında topladığı gözlemlenmiştir. Konu 3 yanlış adrese teslimatı gösteren kelimeler içerirken, konu 16 kargonun yanlış kişiye

teslimatını anlatmaktadır. Görece daha az rastlanan konular ise kayıp kargo ve fatura problemlerini içeren 11'inci konu ve süreklilik arz eden problemleri belirten 9'uncu konu olarak gözlemlenmiştir.

Tablo 5 ise hizmet kalitesi boyutlarına göre toplam şikayet sayılarını göstermektedir. Hizmetin zamanında ve eksiksiz yerine getirilmesini belirten güvenilirlik boyutunun hem konu hem de müşteri şikayeti bazında en yüksek değeri aldığı görülmektedir. Güvenilirlik boyutu altında 8 konu ve toplam şikayetlerin yarısından fazlası toplanmaktadır. Heveslilik boyutu ise şikayetlerin %24'ünü ve dört adet konuyu çatısı altında barındırmıştır. Güven ve empati boyutları ikişer konuya sahipken, sırasıyla şikayetlerin %13 ve %12'si bu iki boyut ile ilgilidir. Dikkat çeken bir bulgu ise fiziksel özellikler hizmet kalitesi boyutunun altında hiçbir konu veya şikayete ulaşamamış olmasıdır.

Tablo 5

Hizmet Kalitesi Boyutlarına Göre Şikayet Sayıları

	Güvenilirlik	Heveslilik	Güven	Empati	Fiziksel Özellikler	Toplam
Şikayet Sayısı %	42.170 %52	19.351 %24	10.273 %13	9.693 %12	0	81487
Konu	(1,3,5,8,10,11,14,15)	(2,4,6,9)	(13,16)	(7,12)		

Müşteri şikayetlerinin temel hizmet kalitesi boyutları altında sınıflandırılmasından sonra firmalar bazında şikayet konuları incelenmiştir. Şekil 4 firmalara göre şikayet konularının sayısını yatay çubuklarla ve yüzdesel dağılımını ısı haritası ile göstermektedir.

KONU	FİRMA_A		FİRMA_M		FİRMA_Y		FİRMA_S		Toplam	
	Şikayet Sayısı	%	Şikayet Sayısı	%	Şikayet Sayısı	%	Şikayet Sayısı	%	Şikayet Sayısı	%
1	1801	%8	1340	%6	972	%5	1916	%10	6029	%7
2	1631	%8	1446	%7	1450	%8	1134	%6	5661	%7
3	1850	%9	1150	%5	1453	%8	672	%4	5125	%6
4	1753	%8	1226	%6	895	%5	1451	%8	5325	%7
5	1295	%6	1199	%6	1176	%6	969	%5	4639	%6
6	1313	%6	1064	%5	742	%4	940	%5	4059	%5
7	1294	%6	1333	%6	1314	%7	1183	%6	5124	%6
8	872	%4	1571	%7	2369	%12	1171	%6	5983	%7
9	1170	%5	1130	%5	961	%5	1046	%6	4307	%5
10	952	%4	1286	%6	1187	%6	1062	%6	4487	%6
11	814	%4	1610	%7	1059	%5	771	%4	4254	%5
12	1251	%6	1229	%6	885	%5	1204	%6	4569	%6
13	1429	%7	1111	%5	1406	%7	1202	%6	5148	%6
14	1381	%6	1606	%7	1066	%6	2153	%11	6206	%8
15	1745	%8	1815	%8	897	%5	989	%5	5446	%7
16	1047	%5	1523	%7	1424	%7	1131	%6	5125	%6

Şekil 4. Firmalara Göre Şikayet Konuları

Buna göre her firmanın farklı konularda şikayet yoğunlukları olduğu görülmüştür. Diğer bir deyişle firmaların ortak bir şikayet konusu bulunmamıştır. A firmasının yüksek şikayet aldığı konular sırasıyla 3, 1, 4, 15 ve 2 olurken en az şikayet aldığı konular 11, 8 ve 10 olarak belirlenmiştir. Y firmasının ise şikayetlerinin 12%'sini 8'inci konu oluştururken, S firmasının şikayetlerinin 26%'sını 1. ve 14. konu oluşturmaktadır. M firmasının ise her konudaki şikayet dağılımları benzerken en yoğun şikayeti 15'inci ve 11'inci konudan aldığı görülmektedir.

Tablo 6
Firmaların Hizmet Kalitesine Göre Şikayet Sayıları

	FİRMA_A		FİRMA_M		FİRMA_Y		FİRMA_S	
	Şikayet Sayısı	%	Şikayet Sayısı	%	Şikayet Sayısı	%	Şikayet Sayısı	%
Güvenilirlik	10710	%50	11577	%54	10179	%53	9703	%51
Heveslilik	5867	%27	4866	%22	4048	%21	4571	%24
Empati	2545	%12	2562	%12	2199	%11	2387	%13
Güven	2476	%11	2634	%12	2830	%15	2333	%12

Tablo 6'ya göre tüm firmalarda şikayetlerin yarısından fazlası güvenilirlik boyutundan gelmiştir. Buna rağmen güvenilirlik boyutundan görece daha düşük şikayet alan A firması olmuştur. Müşterilere karşı hevesli ve yardımsever olma, hizmeti yerine getirme ve sorunları çözme isteğini belirten "heveslilik" ise tüm firmalarda en çok şikayet alan ikinci hizmet kalitesi boyutu olarak karşımıza çıkmıştır. Empati boyutu ve güven boyutunda Y firması hariç diğer tüm firmaların yakın sayıda şikayet aldığı görülmüştür. Ancak Y firmasının çalışanların bilgili, nazik olması, güven duygusu vermesini tanımlayan güven boyutu altında diğer firmalara göre daha fazla şikayet aldığı saptanmıştır.

Sonuç ve Tartışma

Teknolojik gelişmeler, iletişim için fiziksel kanalların zorunluluğunu azaltmış ve mekânsal bağılılığı ortadan kaldırmıştır. Her ne kadar teknoloji pazarları ve iletişimi sanal alana taşıyarak hızlandırmış ve kolaylaştırmış olsa da materyallerin arz noktasından talep notasına ulaşması hala fiziksel dağıtım kanallarıyla mümkün olmaktadır. Bu noktada ise özellikle küçük paket taşınması yapan kargo hizmet sağlayıcılarına olan talep artmıştır.

Kargo firmaları yükselen pazardan pay alabilmek ve rekabette ayakta kalabilmek için müşterilerine üstün hizmet kalitesi sunmak durumundalardır. Bunu sağlayabilmenin en iyi yolu ise sürekli olarak müşteri hizmet kalitesini analiz etmekten geçmektedir. Müşterilerin hizmet deneyimlerini aktardıkları sanal mecralar ise hizmet kalitesinin incelenmesi için büyük miktarda ve farklı tiplerde veri içermektedir. Bu tip büyük verinin analiz edilebilmesi içinse veri madenciliği, makine öğrenmesi ve derin öğrenme gibi metotlar kullanılmaktadır. Araştırmamız Türk kargo sektörünün en büyük dört oyuncusuna yapılan müşteri şikayetlerini toplamış ve metin madenciliği metotlarından biri olan LDA ile analiz etmiştir. LDA, konu modelleme yöntemlerinden biri olup farklı çok sayıda metin içerisindeki konuları ortaya koymaya yarayan bir yöntemdir. Çalışmamız bu metotla kargo firmalarına yapılan şikayetlerin 16 farklı konu altında toplandığını keşfetmiş ve konuları tanımlayan kelimeleri ortaya çıkartmıştır. Gürce ve Tosun (2017)'nin 200 kargo şikayeti içerik analizi inceledikleri çalışmalarında ise 13 konu ortaya konmuştur. Firmalar bazında şikayet konularının dağılımı incelendiğinde araştırmaya dahil edilen firmalarının ortak paylaştığı bir şikayet konusu belirlenmemiştir.

Diğer bir deyişle her bir firmanın karakteristik olarak farklı konular üzerinde şikayetlerinin yoğunlaştığı ortaya çıkmıştır.

En yüksek şikayet alan konu içeriğinin ulaştırma hizmetinin zamanında sağlanamaması olduğu görülmektedir. Araştırmamızda ortaya çıkan bulgu Gürce ve Tosun (2017)'un araştırması ile benzer özellikler taşımaktadır. Aynı zamanda bir diğer sıkça rastlanan şikayet konusu %7.40 ile elektronik ticarete yaşanan kargo sorunlarını kapsamaktadır. Günümüzde önemini ve hacmini giderek artıran e-ticaretin en önemli dağıtım kanalı çözüm ortağının kargo firmaları olması nedeniyle bu konunun müşteri şikayetlerinde ortaya çıkması beklenen bir sonuç olmuştur.

Sonraki adımda ise müşteri şikayetlerinden LDA ile çıkartılan 16 konu araştırmacılar tarafından hizmet kalitesi boyutları olan “fiziksel özellikler”, “empati”, “güven”, “heveslilik” ve “güvenilirlik” altında toplanmıştır.

Fiziksel özellikler hizmet kalitesi boyutu altında bir konuya ve buna bağlı olarak şikayete rastlanmamıştır. Bu durum, kargo hizmet sağlayıcılarıyla müşterilerin fiziksel temasının sınırlı olmasından kaynaklanmış olabilir. Ancak müşteri beklentilerinin doğru bir şekilde anlaşılabilmesi için sonraki çalışmalar için bir araştırma konusu olabilir. Aynı zamanda ortaya koyulan bu sonuç kargo firmalarında fiziksel özellikler boyutunun müşteri memnuniyetine etkisinin anlamlı ancak düşük olduğunu bildiren Onurlubaş ve Gümüş (2020)'ün çalışmasını desteklerken, kargo firmalarında marka sadakatini en çok etkileyen boyutun fiziksel özellikler olduğunu ortaya koyan Karadeniz ve Balcı (2014)'nın araştırması ile çelişmektedir.

Songur (2019)'a göre kargo hizmeti alan müşterilerin memnuniyetini etkileyen en önemli hizmet kalitesi boyutu güvenilirlik olmaktadır. Buna rağmen tüm firmaların şikayetlerinin %50'sinden fazlasının güvenilirlik boyutu altında toplandığı ortaya koyulmuştur. Lojistik yönetiminin temel amacı doğru zamanda, doğru ürünü, doğru miktarda, doğru yere, doğru şekilde teslim etmektir. Müşterilerin lojistik hizmeti veren kargo firmalarından beklentisi de bu yöndedir. Bu nedenle kullanıcı şikayetlerinin yarısından fazlasının “Taahhüt edilen hizmetin kusursuz, güvenilir biçimde yerine getirme becerisini tanımlayan” güvenilirlik boyutundan gelmiş olması beklenen bir sonuç olmuştur.

Yaşanan problemlerin çözümü konusunda firmaların heveslilik boyutu altında ikinci en fazla şikayeti aldıkları görülmektedir. Bu boyut altında toplanan şikayet sayıları, tüm şikayetlerin %25'ini oluşturmaktadır. Ek olarak, güvenilirlik boyutunu sırasıyla, heveslilik, güven, empati izlemiş ve bulgular literatürle örtüşmüştür (Burucuoğlu ve Yazar, 2020; Yıldız, 2020; Gürce ve Tosun, 2017; Çakmak ve Özkan, 2017; Deniz ve Gödekmerdan, 2011; Kayabaşı, 2010).

Pareto ilkesi, birçok sonuç için, sonuçların %80'inin, nedenlerin %20'sinden kaynaklandığını belirtir. Araştırma sonuçları firmalar bazında en çok şikayet aldıkları konuları ortaya koymuştur. Bu durumda yöneticiler çalışmamızın çıktılarında yola çıkarak öncelikle hangi probleme müdahale etmeleri gerektiğini daha iyi analiz edebilirler. Şikayetlerin %80'inin azaltılması için şikayete neden olan %20'lik sorun kaynağının bulunması için araştırmamız sektöre metodolojik ve stratejik bir katkıda bulunmaktadır. Aynı zamanda metin madenciliği, makine öğrenmesi ve derin öğrenme gibi metotların büyük veri ile kullanıldığı taktirde firmalar için önemli stratejik bilgiler yaratma gücüne sahip olduğu ortaya koyulmuştur.

Uluslararası alanda sosyal bilimlerin farklı dallarında gizli dirichlet ayrımı metoduyla müşteri yorumlarının incelendiği çalışmalara rastlamak mümkündür. Ancak Türkçe literatür ele alındığında ise LDA ile yapılan çalışmaların sınırlı olduğu görülmektedir. Türkçe sosyal bilgiler yazınında genel olarak yönetim bilişim sistemleri alanı altında çalışılmış olan bu metodun örneklerine işletme bilim alanında ulaşılamamıştır. Bu

bakımdan arařtırmamız Türkçe iřletme bilimi ve pazarlama bilim dalı yazınına özgün bir katkı saęlamaktadır. Aynı zamanda arařtırmamızda kargo tařımacılıęında çok sayıda müşteri Őikayetinin incelenmiř ve Őikayet konuları, konuları oluřturan kelimelerle keřfedilmiřtir. Bu sonular sonraki alıřmalar için de yol gsterici olmaktadır.

Takip eden alıřmalarda ortaya koyulan Őikayet grupları ve kelimeler kullanılarak bir firmaya gelen Őikayetler ayrıřtırılabilir. Ek olarak zamana gre Őikayet konularının veya sayılarının deęiřimi incelenebilir. Veri seti ve rnekleme firma sayısı veya Őikayetler artırılarak geniřletilebilir. Ek olarak sunulan LDA modeli ile kargo firmaları için otomatik Őikayet sınıflaması yapan bir uygulama geliřtirilebilir.

Arařtırmanın birtakım limitleri bulunmaktadır. ncelikle seilen rnekleme her ne kadar iki yıl (2020-2022) kadar geniř bir zaman dilimini ierse de kısıtlı bir sreci kapsamaktadır. Bununla beraber kargo rnekleme sadece drt adet firmaya ait müşteri Őikayetlerini iermektedir. Ek olarak alıřmada sadece olumsuzluk ieren müşteri Őikayetleri kullanılmıř ve müşteri yorumları tek kaynaktan toplanmıřtır. Son olarak ise kargo sektr iřletmeden iřletmeye (B2B), tketiciden tketicieye (C2C), tketiciden iřletmeye (B2C) ve aynı zamanda iřletmeden tketicieye (B2C) hizmet vererek geniř bir müşteri kitlesini kapsamaktadır. Toplanan Őikayetler arasında verilen hizmete dair bir ayrım yapılamamıřtır.

Kaynaka

- Akın, A. A. ve Akın, M. D. (2007). Trk dilleri iin aık kaynaklı doęal dil iřleme ktphanesi: ZEMBEREK. *Elektrik Mhendislięi*, 431, 38-44. Eriřim adresi: https://www.emo.org.tr/ekler/c7a625d5077d3ba_ek.pdf?dergi=4
- Ala, T. ve Uęuz, S. (2021). Trkiye’de blgesel kalkınmanın giriřimcilik, inovasyon ve Ar-Ge alıřmalarıyla iliřkisinin bibliyometrik analizi ve LDA Mallet uygulaması [Blge Bilimi ve Planlama Kongresi zel Sayısı]. *Erzincan niversitesi Sosyal Bilimler Enstits Dergisi*, 14(20), 13-29. doi: 10.46790/erzisosbil.968808
- Atılgan, K. . ve Yoęurtcu, H. (2021). Kargo firması müşterilerinin Twitter gnderilerinin duygu analizi. *aę niversitesi Sosyal Bilimler Dergisi*, 18(1), 31-39. Eriřim adresi: <https://dergipark.org.tr/tr/pub/cagsbd/issue/63182/959747>
- Aytuę, O., Yalın, A. ve Erdem, A. (2020). niversite bilgi ynetim sistemi servis destek taleplerinin konu modelleme tabanlı analizi. *Avrupa Bilim ve Teknoloji Dergisi*, (Ejosat zel Sayı 2020), 389-397. doi: 10.31590/ejosat.780642
- Barde, B. V. ve Bainwad, M. A. (2017). An overview of topic modeling methods and tools. International Conference on Intelligent Computing and Control Systems Konferansında Sunulan Bildiri, Maduri, Hindistan. Eriřim adresi: <https://ieeexplore.ieee.org/document/8250563/citations?tabFilter=papers#citations>
- Bastani, K., Namavari, H. ve Shaffer, J. (2019). Latent Dirichlet allocation (LDA) for topic modeling of the CFPB consumer complaints. *Expert Systems with Applications*, 127, 256-271. doi: <https://doi.org/10.1016/j.eswa.2019.03.001>
- Blei, D. M. (2012). Probabilistic topic models. *Communications of the ACM*, 55(4), 77-84. doi:10.1145/2133806.2133826

- Blei, D. M., Ng, A. Y. ve Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research*, 3, 993-1022. Erişim adresi: <https://www.jmlr.org/papers/volume3/blei03a/blei03a.pdf?ref=https://githubhelp.com>
- Blei, D. M. ve Mcauliffe, J. D. (2007). Supervised topic models. J. Platt ve diğerleri (Ed.) *Neural Information Processing Systems: Proceedings of Advances in Neural Information Processing Systems 20* (s. 121-128) içinde. Vancouver, Canada. doi: 10.48550/arXiv.1003.0783
- Buenano-Fernandez, D., Gonzalez, M., Gil, D. ve Luján-Mora, S. (2020). Text mining of open-ended questions in self-assessment of university teachers: An LDA topic modeling approach. *IEEE Access*, 8, 35318-35330. doi: 10.1109/ACCESS.2020.2974983
- Burucuoğlu, M. ve Yazar, E. E. (2020). Üçüncü parti platformda kargo firmalarına yapılan müşteri şikayetlerinin içerik analizi. *Ekonomik ve Sosyal Araştırmalar Dergisi*, 16(1), 99-114. Erişim adresi: <https://dergipark.org.tr/tr/pub/esad/issue/54125/645430>
- Büyükkeklik, A., Özoğlu, B. ve Bülbül, H. (2014). Kargo hizmet sağlayıcılarında kalitenin tüketici davranışına etkisi: bireysel tüketici araştırması. *Selçuk Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 31(1), 33-43. Erişim adresi: <https://dergipark.org.tr/tr/pub/susbed/issue/61810/924667>
- Chang, J., Gerrish, S., Wang, C., Boyd-Graber, J. ve Blei, D. (2009). Reading tea leaves: How humans interpret topic models. *Advances in neural information processing systems*, 22. Erişim adresi: <https://proceedings.neurips.cc/paper/2009/file/f92586a25bb3145facd64ab20fd554ff-Paper.pdf>
- Cronin, J. ve Taylor, S. (1992). Measuring service quality: a reexamination and extension. *Journal of Marketing*, 56(3), 55-68. doi: <http://dx.doi.org/10.2307/1252296>
- Çakmak, A. ve Özkan, B. (2017). Kargo kullanıcılarının önem verdikleri faktörlerin, kargo firmaları tarafından başarımlarının incelenmesi. *Tarih, Kültür ve Sanat Araştırmaları Dergisi*, 6(4), 1010-1028. doi: 10.7596/Taksad.V6i4.1036
- Çalli, L., Çalli, F. ve Çalli, B. A. (2021). Yönetim bilişim sistemleri disiplinde hazırlanan lisansüstü tezlerin gizli dirichlet ayrımı algoritmasıyla konu modellemesi. *MANAS Sosyal Araştırmalar Dergisi*, 10(4), 2355-2372. doi: 10.33206/mjss.894809
- Daud, A., Li, J., Zhou, L. ve Muhammad, F. (2010). Knowledge Discovery Through Directed Probabilistic Topic Models: A Survey. *Frontiers Of Computer Science in China*, 4(2), 280-301. doi: 10.1007/s11704-009-0062-y
- Deniz, A. ve Gödekmerdan, L. (2011). Müşterilerin kargo firmalarının sunduğu hizmetlere yönelik tutum ve düşünceleri üzerine bir araştırma. *Atatürk Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, 15(2), 379-396. Erişim adresi: <https://dergipark.org.tr/tr/pub/ataunisobil/issue/2828/38343>
- Türkiye kargo pazarının büyüme için müşteri deneyimini iyileştirmeye odaklanması gerekiyor, 2022. (2022, Kasım). *Devirsaati Ticari Taşıtlar Dergisi*. Erişim adresi: <https://devirsaati.com/turkiye-kargo-pazarinin-buyume-icin-musteri-deneyimini-iyilestirmeye-odaklanmasi-gerekliyor/>
- Eickhoff, M. ve Neuss, N. (2017). *Topic modelling methodology: its use in information systems and other managerial disciplines*. 25th European Conference on Information Systems (ECIS) sunulan bildiri, Guimaraes, Portekiz. Erişim adresi: https://aisel.aisnet.org/ecis2017_rp/86/

- Erman, R. A. (2022). Konu modellemeye dayalı bir izleyici araştırması: survivor Türkiye Yunanistan 2019 (Yüksek Lisans tezi, Hacettepe Üniversitesi, Ankara). Erişim adresi: <http://www.openaccess.hacettepe.edu.tr:8080/xmlui/handle/11655/26221>
- Fabien, L. (2005). Design and implementation of a service guarantee. *Journal of Services Marketing*, 19(1), 33-38. <http://dx.doi.org/10.1108/08876040510579370>
- Garip, S. (2021). Tüketici ekstra rol davranışı ve algılanan hizmet kalitesi arasındaki ilişkiyi tanımlamaya yönelik nitel bir araştırma. *Middle Black Sea Journal of Communication Studies*, 6(1), 46-57. Erişim adresi: <https://dergipark.org.tr/en/download/article-file/1413370>
- Grönroos, C. (1984). A service quality model and its marketing implications. *European Journal of Marketing* 18(4), 36-44. doi: 10.1108/EUM0000000004784
- Gümüş, N. ve Onurlubas, E. (2020). Kargo firmalarının hizmet kalitesinin müşteri memnuniyeti üzerine etkisinin servqual ölçeği ile incelenmesi. *Kırklareli Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 9(1), 30-46. Erişim adresi: <https://dergipark.org.tr/tr/pub/klujfeas/issue/53391/658661>
- Günay, K. ve Güçdemir, Y. (2022). İklim İletişimi Bağlamında 2020-2021 Türkiye’de Sivil Toplum Kuruluşların Twitter Paylaşımlarının Konu Modelleme Analizi. *Turkish Online Journal of Design Art and Communication*, 12(4), 1026-1045. doi: 10.7456/11204100/009
- Gürce, M. Y. ve Tosun, P. (2017). Kargo hizmetlerine ilişkin müşteri şikâyetleri: bir içerik analizi. *İşletme Araştırmaları Dergisi*, 9(3), 177-196. doi: 10.20491/isarder.2017.294
- Jia, S. (2019). Toward a better fitness club: evidence from exerciser online rating and review using latent dirichlet allocation and support vector machine, *International Journal of Market Research*, 61(1), 64-76. doi: 10.1177/1470785318770571
- Kaplan, S. ve Vakili, K. (2015). The double-edged sword of recombination in breakthrough innovation. *Strategic Management Journal*, 36(10), 1435-1457. doi: 10.1002/smj.2294
- Karadeniz, M. ve Balci, M. (2014). Lojistik faaliyetlerde algılanan hizmet kalitesinin marka sadakatine etkisi. *Marmara Üniversitesi İktisadi ve İdari Bilimler Dergisi*, 36(1), 293-315. doi: 10.14780/iibdergi.201417549
- Kayabaşı, A. (2010). Elektronik (online) alışverişte lojistik faaliyetlere yönelik müşteri şikâyetlerinin analizi ve bir alan araştırması. *İşletme Araştırmaları Dergisi*, (2), 21-42. Erişim adresi: <https://www.isarder.org/tr/2010cilt2no2/b22.pdf>
- Kiatkawsin, K., Sutherland, I. ve Kim, J. Y. (2020). A comparative automated text analysis of Airbnb reviews in Hong Kong and Singapore using latent dirichlet allocation. *Sustainability*, 12(16), 6673. doi: 10.3390/su12166673
- Kızılöz, H. E. (2020). Bilimsel makalelerin atif sayısı tahmini. *Avrupa Bilim ve Teknoloji Dergisi, (Ejosat Özel Sayı 2020)*, 370-375. doi: 10.31590/ejosat.araconf48
- Lui, L., Tang, L., Dong, W., Yao, S., Zhou, W. (2016). An overview of topic modeling and its current applications in bioinformatics. *Springerplus.*, 5(1608), 1-22. doi: 10.1186/s40064-016-3252-8
- Mohr, J. W. ve Bogdanov, P. (2013). Introduction—topic models: what they are and why they matter. *Poetics*, 41(6), 545-569. doi: 10.1016/j.poetic.2013.10.001

- Padarian, J., Minasny, B., and McBratney, A. B. (2020). Machine learning and soil sciences: a review aided by machine learning tools. *SOIL*, 6, 35–52, <https://doi.org/10.5194/soil-6-35-2020>.
- Parasuraman, A., Zeithaml, V. A., Berry, L. (1988). SERVQUAL: a multiple-item scale for measuring consumer perceptions of service quality. *Journal of Retailing*, 64(1), 12-40. Erişim adresi: <https://www.sciencedirect.com/journal/journal-of-retailing>
- Parasuraman, A., Zeithaml, V. A. ve Berry, L. L. (1985). A conceptual model of service quality and its implications for future research. *Journal of Marketing*, 49(4), 41-50. doi: 10.2307/1251430
- Poushneh, A. ve Rajabi, R. (2022). can reviews predict reviewers' numerical ratings? the underlying mechanisms of customers' decisions to rate products using latent dirichlet allocation (LDA). *Journal of Consumer Marketing*, 39(2), 230-241. doi: 10.1108/JCM-09-2020-4114
- Putri, I. R. ve Kusumaningrum, R. (2017). Latent dirichlet allocation (LDA) for sentiment analysis toward tourism review in Indonesia. *Journal of Physics: Conference Series*, 801(1), 1-6. doi:10.1088/1742-6596/801/1/012073
- Qiao, Z., Zhang, X., Zhou, M., Wang, G. A. ve Fan, W. (2017). *A domain oriented LDA model for mining product defects from online customer reviews*. 50th Hawaii International Conference on System Sciences Konferansında Sunulan Bildiri, Hawaii, USA. Erişim adresi: <http://hdl.handle.net/10125/41376>
- Razik, M. A., Rozar, N. M., Yusoff, M. N. H., Hassan, H., Ramli, B. ve Zain, R. B. M. (2018). Measuring parcel service quality in university using Servqual model. *Int. J. Sup. Chain. Management*, 7(6), 291. doi: 2050-7399
- Řehůřek, R. ve Sojka, P. (2011). *Gensim-statistical semantics in python*. Erişim adresi: <https://www.fi.muni.cz/usr/sojka/posters/rehurek-sojka-scipy2011.pdf>
- Steyvers, M. ve Griffiths, T. (2007). *Probabilistic topic models*. Landauer, T. K., McNamara, D. S., Dennis, S., Kintsch, W. (Ed.) Handbook of latent semantic analysis içinde (s. 427-448). Newyork, London: Routledge.
- Songur, G. ve Büyükkeklik, A. (2016). Kargo hizmet sağlayıcılarında hizmet kalitesi ve kurumsal müşteri memnuniyeti: Konya İli örneği. *Uluslararası İktisadi ve İdari Bilimler Dergisi*, 2(1), 103-119. Erişim adresi: <https://dergipark.org.tr/tr/pub/uiibd/issue/24165/256355>
- Sutherland, I., Sim, Y., Lee, S. K., Byun, J. ve Kiatkawsin, K. (2020). Topic modeling of online accommodation reviews via latent dirichlet allocation. *Sustainability*, 12(5), 1-15. doi: 10.3390/su12051821
- Tirunillai, S. ve Tellis, G. J. (2014). Mining marketing meaning from online chatter: strategic brand analysis of big data using latent dirichlet allocation. *Journal of Marketing Research*, 51(4), 463-479. doi: <https://doi.org/10.1509/jmr.12.0106>
- Topal, B. ve Şahin, H. (2019). Kurumsal kargo taşımacılığında müşteri memnuniyetinin araştırılması. *Akıllı Ulaşım Sistemleri ve Uygulamaları Dergisi*, 2(2), 14-26. Erişim adresi: <https://dergipark.org.tr/en/download/article-file/833640>
- Toubia, O., Iyengar, G., Bunnell, R. ve Lemaire, A. (2019). Extracting features of entertainment products: a guided latent dirichlet allocation approach informed by the psychology of media consumption. *Journal of Marketing Research*, 56(1), 18-36. doi: 10.1177/0022243718820559

Wang, W., Feng, Y. ve Dai, W. (2018). Topic analysis of online reviews for two competitive products using latent dirichlet allocation. *Electronic Commerce Research and Applications*, 29, 142-156. doi: 10.1016/j.eierap.2018.04.003

Yee, H. L. ve Daud, D. (2011). Measuring customer satisfaction in the parcel service delivery: a pilot study in Malaysia. *Business and Economic Research*, 1(1), 1-10. doi: 10.5296/ber.v1i1.1125

Yıldız, B. (2020). Kargo firmaları hizmet kalitesinin içerik analizi. *Karamanoğlu Mehmetbey Üniversitesi Sosyal ve Ekonomik Araştırmalar Dergisi*, 22(38), 42-57. Erişim adresi: <https://dergipark.org.tr/tr/pub/kmusekad/issue/55242/675159>

Zhai, C. ve Massung, S. (2016). *Text data management and analysis: a practical introduction to information retrieval and text*. New York, California, United States of America: Morgan Claypool Publishers. doi: <https://doi.org/10.1145/2915031>

Extended Abstract

Purpose

Courier companies have to offer superior service quality to their customers in order to get a share from the rising market and to survive in the competition. The best way to achieve this is to constantly analyze the customer service quality. On the other hand, virtual channels where customers convey their service experiences contain large amounts of different types of data for examining service quality. In order to analyze this type of big data, methods such as data mining, machine learning and deep learning are used. The primary purpose of the research is to discover the service quality of the Turkish courier sector through customer complaints. Secondly, it is a methodological contribution to the marketing literature by processing large amounts of textual data with the text mining method. However, another output of the study is that it sheds light on the company's employees and sector managers to understand the market and customer demands.

Design and Methodology

Complaints written by customers, who received service from four cargo companies with the largest market share in Turkey, to www.sikayetvar.com between the years 2020-2022, and information about the complaint were collected on 13.01.2022 with a script created by the authors in the Python software language and stored in the MySQL database. The database includes customer complaints (textual data), complaint dates, usernames, complaint link; It does not contain personal information of users. In order to analyze the textual data in accordance with the purpose of the research, Latent Dirichlet Allocation (LDA), one of the subject modelling methods, was used. As a result of the analysis, after the word clusters that make up the topics were determined, they were classified by the authors under the service quality dimensions. Afterwards, the service quality dimensions for which complaints were received based on companies and holistically were presented and the findings were discussed together with the current literature.

Findings

After the analysis, 16 topics emerged. When the results are considered in detail, it is revealed that the 14th topic corresponds to 7.62% of all complaints and is the most frequently observed complaint. When the words contained in the relevant subject and customer comments are examined, it is seen that the content of the transportation service is not provided on time. At the same time, another frequently encountered complaint

issue emerged as the 1st topic covering delivery problems experienced in electronic commerce with 7.40%. Relatively fewer common issues were observed as the 11th topic, which includes lost packages and invoice problems, and the 9th topic, which indicates persistent problems. Under the service quality dimensions, there was no topic under the physical characteristics service quality dimension and no complaints related to it. However, it is seen that the reliability dimension, which indicates the timely and complete fulfilment of the service, has the highest value based on both subject and customer complaints. Accordingly, while the reliability dimension gathered under 8 topics, more than half of the complaints were gathered under this service quality dimension. The reliability dimension was followed by responsiveness, assurance, and empathy, respectively, and the findings correlated with the literature. Moreover, it has been observed that more than 50% of all companies' complaints are gathered under the reliability dimension. "Responsiveness", which indicates the willingness to be enthusiastic and helpful towards customers, to provide service and to solve problems, is the second most common service quality dimension in all companies. It has been observed that all other companies, except for Y, have received a close number of complaints in the empathy and assurance dimensions. However, it was determined that company Y received more complaints than other companies under the dimension of assurance, which defines the employees to be knowledgeable, courteous and give a sense of trust.

Research Limitations

The research has some limitations. First, although the selected sample includes a wide time period of two years (2020-2022), it covers a limited period. However, the cargo sample includes customer complaints from only four companies. In addition, only negative customer complaints were used in the study and customer comments were collected from a single source. Finally, the cargo sector covers a wide range of customers by serving business to business (B2B), consumer to consumer (C2C), consumer to business (B2C) and business to consumer (B2C). A distinction could not be made between the collected complaints regarding the service provided.

Implications (Theoretical, Practical and Social)

According to the Pareto principle, approximately 80% of most outcomes are caused by 20% of the factors that contribute to those outcomes. The findings of the study showed which problems were the foundation of the majority of the companies' complaints to them. In this scenario, the results of our study provide the managers with the information they need to conduct a more thorough analysis of which issue requires their attention first. Our study makes a contribution to the industry, both methodologically and strategically, by identifying 20% of the causes of customer complaints in order to eliminate 80% of those complaints. When combined with large amounts of data, certain analytical processes, such as text mining, machine learning, and deep learning, have demonstrated the ability to provide businesses with vitally essential strategic information that can help them make key business decisions. In addition, there were neither concerns nor complaints that were included in the dimension of physical qualities. This may have been a consequence of the limited opportunity for clients to engage in face-to-face interaction with the providers of cargo services. On the other hand, this is something that may be researched more in order to get a better understanding of the expectations of customers.

Originality/Value

In the international arena, it is possible to come across studies examining customer comments with the method of Latent Dirichlet Allocation (LDA) in different branches of social sciences. However, when the Turkish literature is considered, it is seen that studies with LDA are limited. The examples of this method, which was generally studied under the field of management information systems in the Turkish social studies literature, could not be reached in the field of business science. In this respect, our research provides an original contribution to the literature of Turkish business science and marketing science. At the same time, in our

research, a large number of customer complaints in cargo transportation were examined and the complaint subjects were discovered with the words that make up the subjects. These results also provide guidance for future studies. Complaints to a company can be separated by using the complaint groups and words revealed in the following studies. In addition, the change in the number or subject of complaints over time can be examined. The data set and sample can be expanded by increasing the number of companies or complaints.

Arařtırmacı Katkısı: Batuhan ÇULLU (%50), Algın OKURSOY (%50).