



Otel Hizmetlerinin Değerlendirilmesinde Gizli Dirichlet Ayrımı ile Analiz: Kastamonu İli Örneği (Analysis with Latent Dirichlet Allocation in the Evaluation of Hotel Services: The Case of Kastamonu)

İbrahim BUDAK^a , *Alptekin SÖKMEN^b 

^a Kastamonu University, Rectorate, Quality Coordinator, Kastamonu/Turkey

^b Ankara Hacı Bayram Veli University, Faculty of Economics and Administrative Sciences, Department of Business Administration, Ankara/Turkey

Makale Geçmişi

Gönderim Tarihi: 26.10.2022

Kabul Tarihi: 16.12.2022

Anahtar Kelimeler

Otel

Gizli Dirichlet ayrımı

Kastamonu

Keywords

Hotel

Latent Dirichlet allocation

Kastamonu

Makalenin Türü

Araştırma Makalesi

Öz

Bu çalışmada, Kastamonu ilinde yer alan otellere yönelik kullanıcıların TripAdvisor sitesi üzerinden yaptığı yorum, görüş ve değerlendirmeler Gizli Dirichlet Ayrımı (GDA) ile analiz edilmiştir. Analiz doğrultusunda elde edilen anahtar kelimeler ile de kümeleme yapılmıştır. Sonrasında kümelere çeşitli başlıklarda etiketleme yapılarak yorumlar görselleştirilmiştir. Yapılan analiz sonucunda en çok İstanbul'daki kullanıcıların Kastamonu'yu ve gene ağırlıklı olarak ağustos ayında ziyaret ettiği ve gene kullanıcıların ailesiyle birlikte seyahat ettiği görülmüştür. Ayrıca kullanıcı yorumları doğrultusunda temel hizmetler, deneyim, konfor ve rahatlık, değer ve eğlence, yapılacak şeyler etiketleri de çıkarılmıştır. Analiz sonucuna göre en yüksek önem derecesine sahip etiket temel hizmetler (0,271) iken; en düşük önem derecesine sahip etiket ise yapılacak şeyler (0,164) olarak tespit edilmiştir. En çok öne çıkan kelimeler ise sırasıyla otel, oda/odalar, kahvaltı, kibarlık, personel, Kastamonu ve temizliktir.

Abstract

In this study, the comments, opinions and evaluations made by the users on the TripAdvisor site regarding the hotels in Kastamonu province were analyzed with the Latent Dirichlet Allocation (LDA). Clustering was also done with the keywords obtained in line with the analysis. Afterwards, the clusters were labeled under various headings and the comments were visualized. As a result of the analysis, it was seen that the users in Istanbul mostly visited Kastamonu and again in August, and again the users traveled with their families. In addition, based on user comments, the labels of basic services, experience, comfort and convenience, value and fun, and things to do were also removed. According to the results of the analysis, the tag with the highest degree of importance was basic services (0.271); The label with the lowest degree of importance was determined as things to do (0.164). The most prominent words are hotel, room/rooms, breakfast, politeness, staff, Kastamonu and cleaning.

* Sorumlu Yazar

E-posta: alptekinsokmen@gmail.com (A. Sökmen)

DOI: 10.21325/jotags.2022.1124

GİRİŞ

Son zamanlarda, web 'de çok büyük miktarda içerik oluşturulmaktadır. Müşterilerin çoğu, görüşlerini veya puanlarını inceleme sitelerinde yayınlamaktadır. Yönetimsel perspektifte, insanların ne düşündüğünü anlamak için bu verilerin incelenmesi gerekmektedir. Ancak, bu tür çok sayıda veriyi analiz etmek, filtreleme ve sınıflandırma yöntemleri olmadan zaman alıcı ve zahmetlidir. Tüm metin verilerini salt metin tabanlı arabirimle göstermek ve/veya tanımlamak yetersiz kalmaktadır. İnsan bilişinin performansını artırmak, etkili bir sınıflandırma uygulamak ve çıkarılan bilgileri görselleştirme yöntemleriyle temsil etmek önemlidir. Kullanıcılar tarafından yapılan incelemeleri sınıflandırmak için en önemli özelliklerden biri etiketlendirmedir. GDA, etiketleri gizli değişkenlere göre çıkarmak için üretken bir modeldir. Etiketler, korpusu belirli özelliklerini temsil eden kelime torbaları olarak temsil edilmektedir. GDA, kelime ağırlıklarının olasılık dağılımına göre uygun etiketlere göre tahsis etmektedir (Chen vd., 2015: s. 687).

GDA, bir işletmede artan rekabet ve genişleyen satış yelpazesini göstermek, ayrıca şirketin değişken piyasalarda finansal kapasitesini korumak ve/veya geliştirmek için işle ilgili tüm yönleri yeniden inceleme, derleme, düzenleme ve iyileştirme ihtiyacını karşılamaktadır. Etkin bir şekilde yönetilen müşteri ilişkileri yoluyla geliri artırma yeteneği, işletmenin amaçlanan hedeflerine ulaşması için önemlidir. Günümüzde müşteriler web sitesi aracılığıyla işletmelerle kolay bir şekilde iletişim kurabilmektedirler. İnternet teknolojisi aracılığıyla belirli ürün veya hizmetlerin kullanımı ve tanımıyla ilgili gayri resmi iletişim, hizmet sağlayıcılar ile tüketiciler arasında en etkili iletişim olarak kabul edilmektedir (Priyantina & Sarno, 2019: s. 142).

Bu çalışmada, otel hizmetleri ve ürünlerine ilişkin müşteri yorumları, otel yönetimi değerlendirmesi ve iyileştirmesi için bir faktör olmanın yanı sıra, otelin ürün veya hizmetlerindeki hata ve eksiklikleri düzeltmek için, bir tetikleyici faktör olarak GDA ile analiz edilmiştir. Otellerin web sitelerinde yazılan müşteri yorumları, uygun teknikler kullanılarak analiz edilebilmeleri için büyük ve yapılandırılmamış veriler kullanılmıştır. Sonraki aşamada ise, veriler doğrultusunda etiketler oluşturularak görselleştirilmiştir. Çalışmada uygulanan teknik ile bulunan etiketler ve anahtar kelimeler Kastamonu'daki oteller üzerinde ilk olma özelliğini taşımaktadır.

Literatür Taraması

Otel çevrimiçi değerlendirmelerinin metin içi özellik çıkarımı ve konu modelleme yöntemlerinde kullanıldığı görülmektedir. Bu bölümde otellerin çevrimiçi değerlendirmelerini GDA yöntemi ile analiz eden çalışmaların bir kısmı özetlenmiştir.

Afaq ve arkadaşları (2022), Güney Asya'daki beş küresel zincir otelin 94.200 çevrimiçi yorum üzerinde konu modellemesi ve duygu analizi yapmışlardır. Çalışmada veriler web kazıma yoluyla TripAdvisor'dan alınmıştır. Konu modelleme müşteriler tarafından yorumlarda belirtilen temel konuları analiz etmek için veri seti üzerinden GDA analizi yapılmıştır. Analiz sonucuna göre temizlik, yemek, personel ve hizmetin otel misafirlerinin ana endişeleri olduğu belirlenmiştir. Ayrıca, otel misafirlerini etkileyen ana sorunların servis gecikmeleri olduğu belirlenmiştir.

Annisa ve arkadaşları (2019), çalışmalarında Travekola adlı çevrimiçi seyahat acentesinde yer alan otel müşterilerinin yorumlarını GDA kullanarak analiz etmişlerdir. Çalışmada, GDA ile konular belirlenmiş ve müşteriler tarafından sıklıkla kullanılan anahtar kelimeler ortaya çıkarılmıştır. Çalışmanın bulgularında, anahtar kelimelerden sekiz konu (oda memnuniyeti, tesis, kahvaltı, otelin sahile yakın konumu, otel çevresinde turist deneyimi, atmosfer,

ağırlama, banyo) çıkarılmıştır. Ayrıca çalışmanın çıktıları kullanılarak otel yönetimi ve müşteriler arasında bir geri bildirim oluşturması hedeflenmiştir.

Zhang ve arkadaşları (2019), çalışmasında otel hizmetlerinin değerlendirilip iyileştirilmesi amacıyla otel hizmetlerinden yararlanan müşterilerin yorumları analiz edilmiştir. Çalışma, New York City’de bulunan oteller üzerinden gerçekleştirilmiştir. TripAdvisor’daki 27.864 otel yorumları analiz edilmiştir. Yorumlarda tartışılan konuları analiz etmek için GDA’ya benzer STM adlı yeni bir yöntem kullanılmıştır. Yorumlar olumlu ve olumsuz olarak iki şekilde analiz edilmiştir. STM, en olumsuz on konuyu (ciddi hizmet arızası, rezervasyon ve iptal, oda tipi, kirlilik, fazla ücretlendirme, oda olanakları, gürültü, hatalar, deneyim karşılaştırması ve genel tesisler) olarak belirlemiştir.

Akhtar ve arkadaşları (2017), otel rezervasyonu yaparken otel web sitelerinin bu süreçte önemli bir rol oynadığı ancak sitede yer alan tüm yorumların okunması ve bir değerlendirmeye gidilmesinin zor olduğuna vurgu yapmışlardır. Buradaki problemden yola çıkarak yapmış oldukları çalışmada otel yorumları analiz edilmiş ve yorumlarda gözden kaçabilen bilgiler ortaya çıkartılmıştır. Yorumlar ilgili otelin web sitesinden taranmış olup bazı ortak yönler göre önceden tanımlanmış gruplarda sınıflandırılmıştır. Sonrasında ise gizli bilgileri ve yönleri belirlemek amacıyla GDA uygulanmıştır.

Xiang ve arkadaşları (2017), yapmış oldukları çalışmada TripAdvisor, Expedia ve Yelp isimli seyahat sitelerindeki otellere bırakılan yorumlar GDA tekniği kullanılarak analiz edilmiştir. Yapılan analizler karşılaştırmalı bir şekilde incelenmiştir. Yapılan GDA analizi sonucunda yorumlar beş başlık (temel hizmet, değer, görülecek/gezilecek yerler, yemek/ tecrübe etmek, temel ürün) altında değerlendirilmiştir. Bulgular, otel endüstrisinin bu platformlarda temsil edilmesinde büyük farklılıklar olduğunu göstermiştir.

Ekinci ve Omurca (2017), yapmış oldukları çalışmada Türkçe otel yorumlarını ele alarak konu modellemesi yapılmasını ve ürün özelliklerinin çıkarılmasını hedeflemiştir. Çalışmanın veri setini oluşturmak için www.otelpuan.com'dan yararlanılmıştır. Toplamda 1000 yorum GDA tekniği kullanılarak analiz edilmiş ve beş konu başlığında (yiyecek/içecek, imkân, bina, personel, ilgi/alaka) toplanmıştır.

Nguyen ve Ho (2021), otel hizmetlerindeki müşteri deneyimlerinden yararlanarak GDA analizi yapmıştır. Çalışmada ilk olarak 99.322 İngilizce müşteri yorumu derlenmiştir. Bu derlemeden, Şaşırma ve Tutarlılık Skoru ölçümleri kullanılarak en iyi konu sayısı (K) seçilmiştir. Konuların belirlenmesi için K katsayılı GDA konu modeli kullanılmıştır. Çalışmanın bulgularında kullanıcılar için ilginç olan, gizli konular ve anahtar kelimelere ulaşılmıştır. Toplamda on dört konu belirlenmiştir. Çalışma, kullanıcı yorumlarını sınıflandırmak için bir örnek oluşturmuştur.

Guo ve arkadaşları (2017) ise çalışmalarında otel müşterilerinin bıraktığı 266.544 yorumu GDA tekniği kullanarak analiz etmişlerdir. Analiz sonucunda müşterilerin yorumlarından GDA ile otuz konu başlığı çıkarılmıştır. Her konu başlığının içinde anahtar kelimeler ağırlıklarına göre sıralanmıştır.

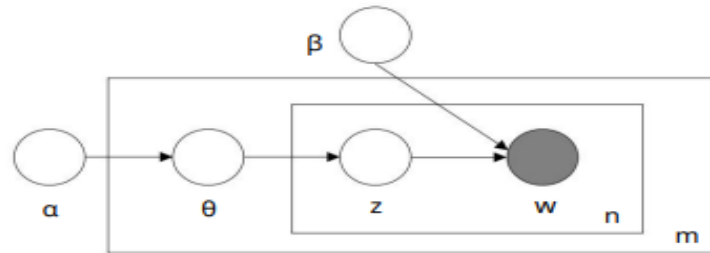
Konaklama hizmetlerinde müşteri memnuniyeti önemlidir. Müşterilerin seyahat hizmetleri sunan çeşitli web sitelerine yapmış oldukları çevrimiçi değerlendirmeler bir sonraki müşterinin konaklama, tercih vs. kararını etkileyebilmektedir. Ayrıca bu değerlendirmeler otel hizmetleri için de bir geri bildirim oluşturmaktadır. Bu alanda yapılmış çalışmalar incelendiğinde genellikle ortak amaç kullanıcıların daha kolay ve etkin bir şekilde ilgili otelle ilgili bir değerlendirmeye ulaşabilmesini sağlama şeklindedir. Ayrıca değerlendirmelerden çıkan konu başlıkları ve

anahtar kelimeler de müşteri memnuniyetinin ve şikayetlerinin belirlenmesinde önemli bir rol oynadığı görülmektedir. Çevrimiçi değerlendirmelerde hangi konuların öne çıktığının belirlenmesinde GDA tekniği literatürde sıklıkla kullanıldığı görülmektedir.

Gizli Dirichlet Ayrımı

Gizli Dirichlet Ayrımı (GDA), olasılık tabanlı denetimsiz bir konu modelleme tekniği olup 2003 yılında Blei ve arkadaşları tarafından tanıtılmıştır (Blei vd., 2003: s. 996). GDA'da geçen gizli kelimesi ile ifade edilmek istenen dokümanlardaki gizli konuların keşfedilmesidir (Jadhav, 2014: s. 1). Modellemenin temeli, belgelerin gizli konular üzerinde rastgele karışımlar olarak temsil edilmesidir. Bu modelleme ile belirlenen her bir konu, kelimeler üzerinden bir dağılımla karakterize edilmektedir. Kısaca, belgelerde yer alan kelime ve kelime ağırlıkları kullanılarak belgelerden konu veya konuların belirlenmesini sağlamaktadır (Blei vd., 2003: s. 996). GDA, kelimeler ve ağırlıklarını kullanarak belgelerdeki trendlerin çözülmesine yardımcı olmaktadır. Belirlenen konu ve kelime dağılımları, o konuyu en iyi şekilde açıklayabilecek kelimelerin tanımlanması ile oluşmaktadır (Bolelli vd., 2009: s. 776-780). GDA algoritması, konu sınıflandırma probleminde sahip olduğu özellikler nedeniyle en çok tercih edilen modellerden birisidir (Nguyen & Ho, 2021: s. 588).

Şekil 1'de GDA grafik modeli verilmiştir. Şekilde yer alan rastgele değişkenler düğümler ile temsil edilmiştir. Kutular, kopya edilmiş plakaları temsil etmektedir. Belgeleri dış plaka temsil ederken bir belge içindeki konuların ve kelimelerin tekrarlanan seçimlerini ise iç plaka temsil etmektedir (Blei vd., 2003: s. 1003-1004).



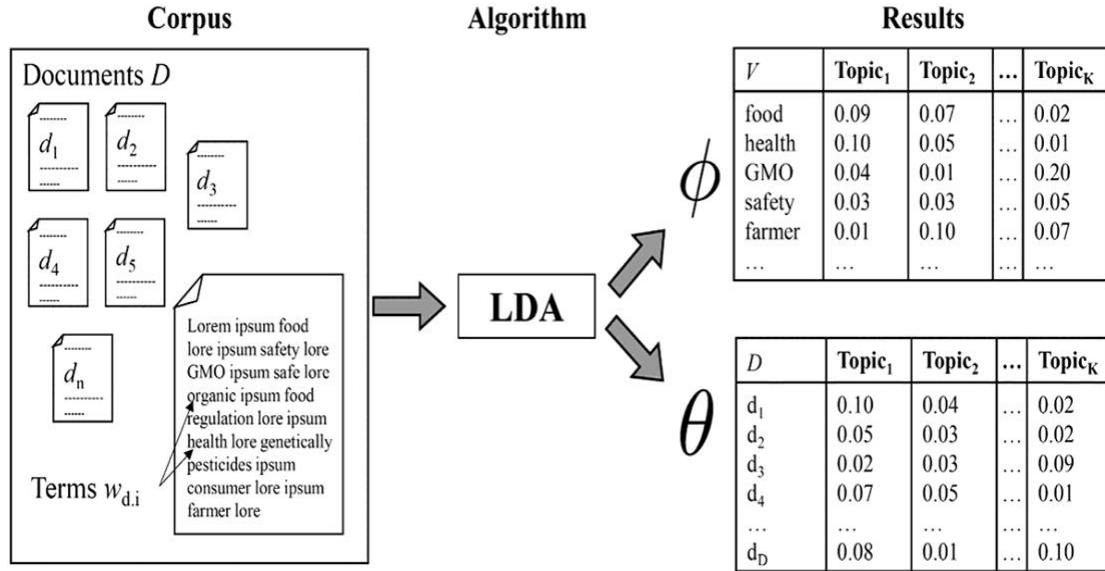
Şekil 1. LDA Grafik Modeli (Blei, Ng & Jordan, 2003: s.997).

Bu modelin formülü şu şekilde verilmiştir:

$$p(\theta, z, w | \alpha, \beta) = p(\theta | \alpha) \prod_{n=1}^N p(z_n | \theta) p(w_n | z_n, \beta)$$

Şekil 1'de verilen modelde; m parametresi, üretilecek belge sayısını; α , konu dağılımını; z , bir konuyu; θ , belirli doküman için konu dağılımını; w , kelimeleri temsil etmektedir. Şekilden de anlaşılacağı üzere, GDA gösteriminin üç seviyesi vardır. Bir derlem oluşturma sürecinde α ve β parametreleri bir kez örneklendiği varsayılmaktadır. θ parametresi ise her bir doküman için ayrı örneklenmektedir (Blei vd., 2003: s. 1003-1004).

GDA, modellenirken dokümanların yer aldığı evren carpus olarak ifade edilirken bu dokümanlar içerisinde yer alan terimler ise sözcükleri ifade etmektedir (Maier vd., 2018: s. 95). Şekil 2'de örnek uygulama modeli verilmiştir.




Şekil 2. LDA Uygulama Örneği (Maier vd.2018: s. 94)

Şekil 2’de verilen GDA uygulamasından görüldüğü üzere dokümanlarda yer alan kelimeler ve ağırlıkları kullanılarak gizli bilgiler ortaya çıkarılmakta ve bu kapsamda konular tespit edilebilmektedir.

Uygulama

Çalışmada, 2000 yılında kurulan, dijital ortamda kullanıcıların hizmetine veren sosyal medya sitesi olan ve şu anda dünyanın en büyük seyahat birliği olarak kabul edilen TripAdvisor sitesi verileri (metinleri) kullanılmıştır (www.tripadvisor.com). Bu çerçevede TripAdvisor’da yer alan Kastamonu’daki otellerin müşteri değerlendirmeleri ele alınmıştır. Aşağıda Şekil 3’de TripAdvisor’da bir müşteriye ait otel için kriter değerlendirilmesi verilmiştir.



Emirhan K.

şu tarihte bir yorum yazdı:

Haz 2022

...

●●●●●

Genel skor

Otel deneyimi

“Otel on numaraydı . Aile ile de kalmaya müsait nezih bir yer. Harika deneyimler yaşadım . Atv surdum ata bindim yemekler lezzetli mükemmel bir doğa ortamının içinde çok sakin huzurlu bir yer . Herkese tavsiye ederim”

[Daha azını göster](#) ▲

Konaklama tarihi: Eylül 2021

Seyahat türü: Arkadaşlarla seyahat etti

Şekil 3. Bir Müşteriye Ait Otel Değerlendirmesi

Şekil 3’de TripAdvisor’da bir otele yapılan değerlendirmede kullanıcı adı, tarih, genel skor, yorum başlığı, yorum, konaklama tarihi ve seyahat türü verilmektedir. Bu kriterler dışında lokasyon, katkı ve faydalı oy kriterleri de kullanıcı isteği doğrultusunda gösterilebilmektedir.

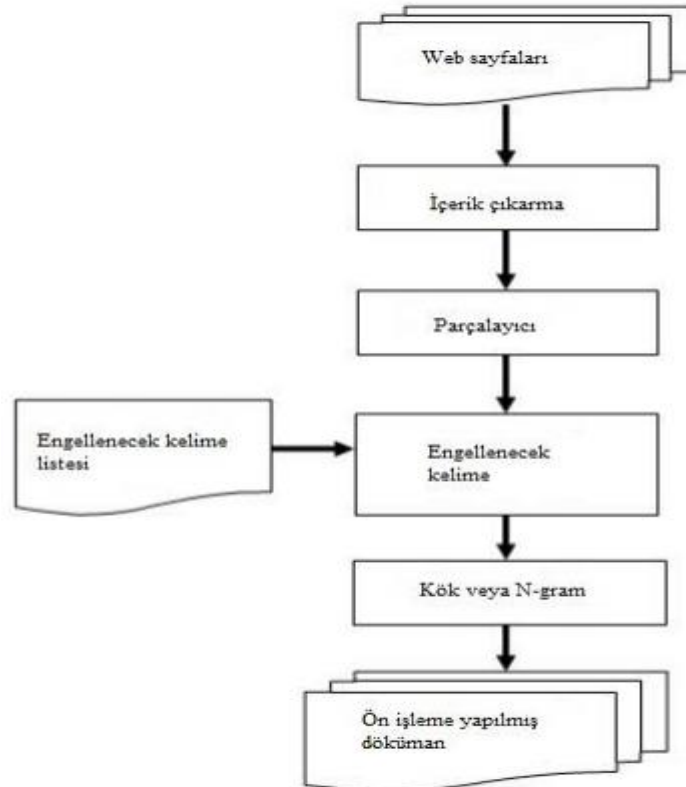
Çalışmada 2022 ikinci çeyreğine kadar yapılan toplam 1687 müşteri değerlendirmesini analizi için açık kaynak kodlu RapidMiner 9.10.1 uygulaması kullanılmıştır. RapidMiner, kullanıcının veri analizi işlemlerini operatörleri tarafından ‘tak-çalıştır’ tarzında tasarlamasına olanak tanıyan ve kod kullanılmasına gerek olmadan çalışan bir veri analizi platformudur. Ayrıca, RapidMiner Marketplace’te sunulan uzantılar geliştirilerek RapidMiner’e işlevsellik de eklenebilir (www.rapidminer.com). Değerlendirmeler, uygulamada yer alan operatörler yardımıyla özellik seçimi yapılarak analiz edilmiştir. Müşteri yorumları dışında yapılan analizde Tablo 1’deki değerlendirme yapılmıştır.

Tablo 1. En sık ziyaret edilen Lokasyon, Konaklama Zamanı ve Seyahat Türünün İncelenmesi

Lokasyon	Kullanıcı Sayısı	Konaklama Zamanı	Kullanıcı Sayısı	Seyahat Türü	Kullanıcı Sayısı
İstanbul	400	Ağustos	250	Ailesiyle seyahat etti	646
Ankara	249	Temmuz	180	İş için seyahat etti	328
Kastamonu	49	Ekim	150	Arkadaşlarıyla seyahat etti	296
Samsun	28	Şubat	143	Çift olarak seyahat etti	197
İzmir	24	Eylül	134	Tek başına seyahat etti	43

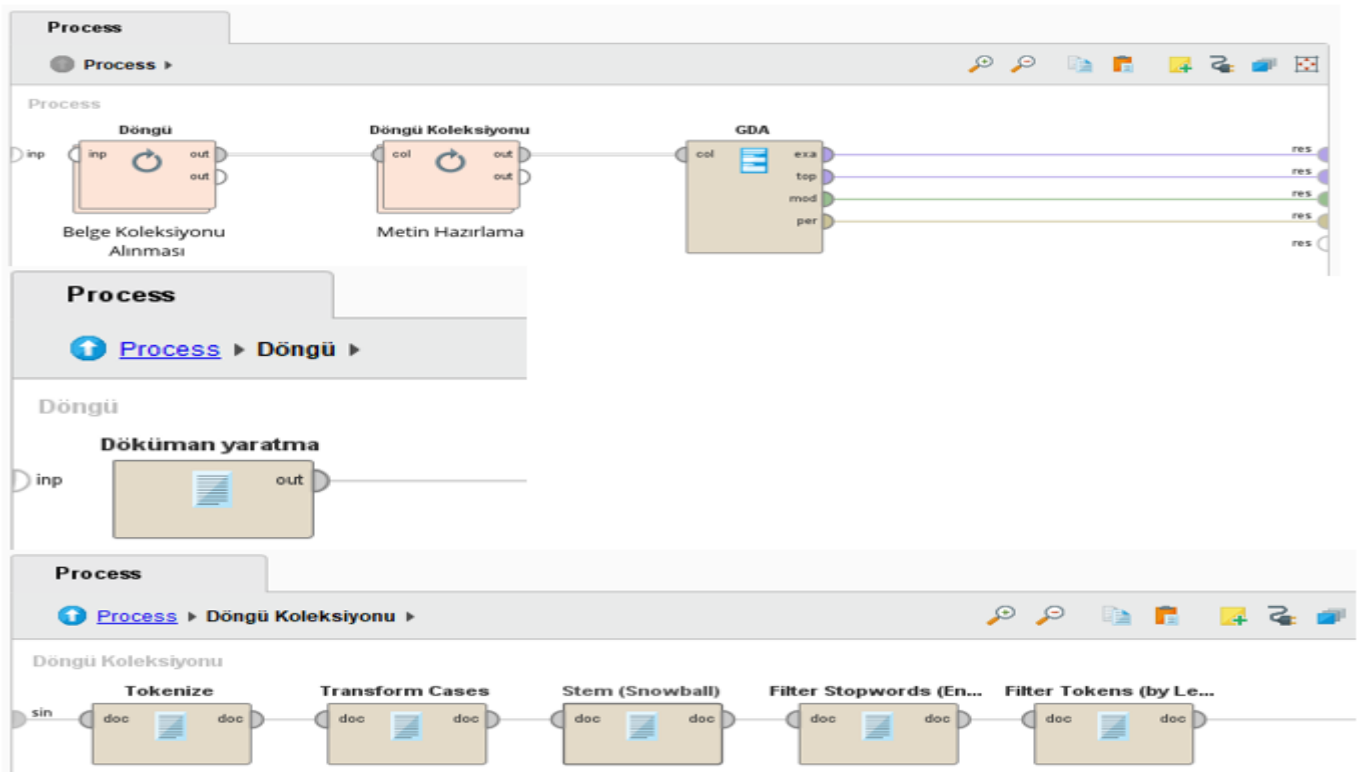
Tablo 1’de yer alan otel değerlendirmeleri kapsamında sık ziyaret edilen lokasyon, konaklama zamanı ve seyahat türü verilmiştir. Lokasyon olarak Türkiye’de bulunan illerin yanı sıra Almanya, Çin, Güney Afrika, Avustralya gibi ülkelerin de bulunduğu 154 farklı noktadan Kastamonu otelleri ziyaret edilmiştir. Ziyaret eden müşteriler yılın farklı zamanlarında gelmekle birlikte Kastamonu’yu en sık ‘ağustos’ ayında ziyaret ederken; en az ziyaret ‘aralık’ ayında gerçekleşmiştir. Gelen müşterilerle ilgili veriler değerlendirildiğinde genel olarak en sık ‘ailesi ile’ ziyaret edildiği buna karşın en az ‘tek başına’ ziyaret edildiği ortaya çıkmıştır.

Çalışmada müşteri yorumları RapidMiner uygulamasında yer alan GDA operatörü doğrultusunda analiz edilmiştir. Analiz edilmeden önce ilgili yorumlar, ön işlem sürecinden geçirilmiştir. Şekil 4’te müşteri yorumlarının ön işleme adımları gösterilmiştir (Budak, 2021: s. 59).



Şekil 4. Ön İşleme Sürecinin Adımları

Şekil 4’de müşteri yorumları web sayfalardan alınarak içerik çıkarımı yapılmış ve sonrasında cümleler parçalanarak engellenecek kelime listesi çıkarılmıştır. Ardından kelimelerin kök kelimelerine veya kelimelerin birlikte anlamlı olacağı n-gram yapılarına ulaşılmıştır. Bu kapsamda kelimenin farklı morfolojik değişimi temel kelimeye eşlemek için gerçekleştirilmiştir. Olasılık hesaplamasına dayalı olarak performans değerlendirmesi yapan Bayes Algoritması tabanlı GDA ile yorumlar anahtar kelimeler kullanılarak etiketlenmiştir. RapidMiner’da yer alan ‘Operator Toolbox’ uzantısı kullanılmıştır. Bu uzantıda Şekil 5’de yer alan Loop (döngü), Create Document (dökmüman yaratma), Loop Collection (döngü koleksiyonu), Tokenize (parçalayıcı), Filter Stopwords (engellenecek kelimeleri filtreleme), Transform Cases (metin dönüşümü), Stem (kelime kökü), Filter Length (kelime uzunluğu) ve GDA operatörleri ile analiz edilmiştir.



Şekil 5. Operator Toolbox Uzantısı ile GDA Modeli

Şekil 5’de yer alan GDA operatörü ile uygun etiket ve anahtar kelime sayısını ayırt etmek için farklı iterasyonlarda hesaplama yapılarak çeşitli varyasyonlar denenmiştir. Gerçekleştirilen hesaplamada 1000 iterasyonda 5 etiket ve her bir etiket oluşumunda 10 anahtar kelime olacak şekilde düzenleme yapılmıştır. Aşağıda Tablo 2’de 1687 yorumdan ön işlenmiş kelime listesi oluşturularak GDA ile analiz edilen etiketler ve anahtar kelimeler verilmiştir.

Tablo 2. GDA ile Etiket ve Anahtar Kelimelerin Belirlenmesi

Değer ve Eğlence		Konfor ve Rahatlık		Deneyim		Yapılacak Şeyler		Temel Hizmetler	
Kelime	Ağırlık	Kelime	Ağırlık	Kelime	Ağırlık	Kelime	Ağırlık	Kelime	Ağırlık
Personel	2580	Odalar	3605	Otel	4455	Tabiat gezisi	1360	Otel	5880
Tavsiye	1265	Temizlik	2265	Yer	1660	Şehir gezisi	985	Oda	3265
Çok güzel	1240	Merkezi	1015	Konaklama	1645	Tatil	915	Kahvaltı	3020
Yemekler	1025	Konum	975	Geçirilen zaman	1400	Teşekkür	905	Kibarlık	2805
Bahçe ve çevre	815	Olanaklar	940	Güzellik	1315	Aklı almak	810	Kastamonu	2455
Aileye uygun	800	Şükran	860	Konak	1210	İş çalışmaları	770	Destek	1860
Gerçekleşenler	760	Ev sahipleri	795	Rahatlık	1160	Fiyatlama	740	Yararlı	1775
Tarihi değer	735	Bahsedilen	685	Restoran	1085	Ferah	690	Hizmetler	1660
Arkadaş ortamı	675	Nefis	665	Ulaşım	980	Akşam	680	Konaklama	1455
Yapılanlar	655	Eksiksiz	635	Yemekler	775	Bölge gezisi	540	Çalışanlar	1280

Tablo 2’de GDA modeli ile yorumlar 5 etiket altında analiz edilmiş ve en sık taranan anahtar kelimeler kümelerine ayrılmıştır. Her bir etiket altında sık rastlanan kelimeler bazında araştırmacı ve uzman tarafından değerlendirilerek başlık verilmiştir. Değer ve Eğlence etiketinde sık taranan personel, tavsiye, çok güzel; Konfor ve Rahatlık etiketinde odalar, temizlik, merkezi; Deneyim etiketinde otel, yer, konaklama; Yapılacak Şeyler etiketinde tabiat gezisi, şehir gezisi, tatil; Temel Hizmetler etiketinde otel, oda, kahvaltı kelimeleri yer almaktadır. Şekil 6’da her bir etikette yer alan anahtar kelimeler kullanım sıklığına göre görselleştirilmiştir.

**Şekil 6.** Etiketler ve Kelime Bulutları

Şekil 6’da yorumların beş farklı etiket altında en sık yer alan popüler kelime bulutları verilmiştir. Tablo 3’de GDA modeli ile yorumların etiketlere göre dağılımı verilmiştir.

Tablo 3. GDA Modeli ile Yorumların Etiketlere Göre Dağılımı ve Tahmini

Tahmin	Temel Hizmetler	Deneyim	Konfor ve Rahatlık	Değer ve Eğlence	Yapılacak Şeyler
Temel Hizmetler	0,271	0,212	0,183	0,169	0,164

Tablo 3’de GDA modeli ile yapılan yorumların etiket dağılımı ve tahmini verilmiştir. Bu dağılımda sırasıyla temel hizmetler ‘0,271’; deneyim ‘0,212’; konfor ve rahatlık ‘0,183’; değer ve eğlence ‘0,169’; yapılacak şeyler ‘0,164’ değerler ortaya çıkmıştır. Bu değerlere göre temel hizmetlerin en çok, yapılacak şeylerin ise en az dağılıma sahip olduğu belirlenmiştir. Bu doğrultuda bir sonraki yazılacak yorumun % 27,10 ihtimalle temel hizmetler etiketi altında

yer alan anahtar kelimelerin gelmesi öngörülmektedir. GDA modeli sonucunda α değeri 14,297 iken, β değeri ise 354,150'dir.

Sonuç ve Öneriler

Kastamonu otellerinin müşteri değerlendirilmesi yapılırken öne çıkan etiketler / konular / temalar GDA analizi ile gerçekleştirilmiştir. Analiz sonucunda 5 farklı etiket ve her bir etiket altında yer alacak anahtar kelimeler çıkarılmıştır. Anahtar kelimeler ile sözcük ilişkilendirmesi yapılarak öncelikle etiketler isimlendirilmiştir. Etiketler içerisinde yer alan anahtar kelimeler uzman ve araştırmacılar tarafından değerlendirilerek en uygun şekilde başlıklara ayrılmıştır. Bu doğrultuda etiketlerin hangi eğilimde ilerleyeceği de nicel olarak hesaplanmıştır. Kastamonu otelleri için yapılan 1687 yorumun etiketleri şu şekildedir:

- Etiket 1: Temel Hizmetler
- Etiket 2: Deneyim
- Etiket 3: Konfor ve Rahatlık
- Etiket 4: Değer ve Eğlence
- Etiket 5: Yapılacak Şeyler

Buna göre Kastamonu otelleri bağlamında ele alınan yorumların bu beş temel konu üzerinde yoğunlaştığı söylenebilir. Makalelerin konu dağılımı ve tahminlerine bakıldığında Etiket 1'de '0,271'; Etiket 2'de '0,212'; Etiket 3'te '0,183'; Etiket 4'te '0,169'; Etiket 5'te '0,164' değerleri ortaya çıkmıştır. Bu değerler sonucunda bir sonraki makalelerin daha çok Etiket 1'den (Temel hizmetler) gelmesi öngörülmektedir.

Kastamonu otellerinin yorum değerlendirmesi yanı sıra lokasyon, konaklama zamanı ve seyahat türü de analiz edilmiştir. Lokasyon olarak dünyanın 154 farklı noktasından turist ve ziyaretçi ile karşılaşmıştır. Türkiye özelinde ise sırasıyla en fazla İstanbul, Ankara, Kastamonu, Samsun ve İzmir'den turist ve ziyaretçilerin geldiği görülmektedir. Gelen ziyaretçiler çoğunlukla ağustos, temmuz, ekim, şubat ve eylül aylarını tercih etmişlerdir. Seyahat türü dağılımında ise sırasıyla aile, iş, arkadaş, çift ve tek başına ziyaretlerin gerçekleştirildiği görülmüştür.

GDA ile yapılan bu çalışmada, TripAdvisor sitesinde yapılan kullanıcı değerlendirmelerinin ne doğrultuda eğilim gösterdiğinin ortaya konmasına çalışılmıştır. GDA analizi ile farklı coğrafi yerlere de analizler yapılarak karşılaştırma yapılabilmesi mümkündür. Yapılan analizler doğrultusunda gerek merkezi ve yerel yönetim, gerekse de Turizm işletmeleri tarafından uygun politika ve çalışmalar gerçekleştirilebilir. Bu kapsamda gerçekleşen çalışmada GDA analizi ile Kastamonu turizmüne yönelik eğilimlerin belirlenmesi ve buna uygun orta ve uzun vadeli stratejik karar/kararlar alınması noktasında önemli katkı sağlanabilecektir. Analizde ortaya çıkan etiketlerin/konuların/temaların önem dereceleri belirlenerek etkin çıkarımlar yapılabilir. Çıkan etiketler altındaki anahtar kelimeler ile ulusal ve uluslararası bağlamdaki müşteri portföyü için geliştirmeler rahatlıkla yapılabilir. Örneğin, yakın lokasyon olmasına rağmen Ankara ağırlıklı talep yaratılması için kampanya ve teşvik uygulamalarına gidilebilir. Buna ek olarak bu çalışmada öne çıkan ve önem derecesi en yüksek olan 'temel hizmetler' etiketine yoğunlaşarak araştırma yapılabilir. Detaylandırmak gerekirse, Temel Hizmetler Etiketi altında yer alan otel seçimi, oda özellikleri, kahvaltı zenginliği ve çeşitliliği, gösterilen nezaket ve kibarlık, sunulan hizmetler ve çalışan özellikleri, turist ve ziyaretçiler için en önemli unsurları oluşturmaktadır. Bu kapsamda Kastamonu genelinde konaklama imkânı sunan işletmelerin öncelikle bu unsurlara yoğunlaşarak geliştirmeleri hem memnuniyeti, hem

konaklama sürelerini, hem de tekrar aynı işletmelerin seçimini etkileyebilecektir. Bunun tamamlayıcısı olarak da, yerel yönetim ve Kastamonu Üniversitesinin de bu doğrultuda geliştirici faaliyet ve eğitim desteği sunmaları beklenebilir.

Beyan

Makalenin tüm yazarlarının makale sürecine verdikleri katkı eşittir. Yazarların bildirmesi gereken herhangi bir çıkar çatışması yoktur.

KAYNAKÇA

- Afaq, A., Gaur, L., & Singh, G. (2022). A latent dirichlet allocation technique for opinion mining of online reviews of global chain hotels. *In 2022 3rd International Conference on Intelligent Engineering and Management (ICIEM)*, 201-206.
- Akhtar, N., Zubair, N., Kumar, A., & Ahmad, T. (2017). Aspect based sentiment oriented summarization of hotel reviews. *Procedia Computer Science*, 115, 563-571.
- Annisa, R., & Surjandari, I. (2019). Opinion mining on mandalika hotel reviews using latent dirichlet allocation. *Procedia Computer Science*, 161, 739-746.
- Blei, M. D., Ng, A. Y.-T., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research*, 993-1022.
- Bolelli, L., Ertekin, Ş., & Giles, C. L. (2009). Topic and trend detection in text collections using latent dirichlet allocation. *Proceedings of the 31th European Conference on IR Research on Advances in Information Retrieval*. Toulouse, 776-780.
- Budak, İ. (2021), *Veri ve Metin Madenciliği ile Hava Yolu İşletmelerinin Sosyal Medya Yorum Ve Skorlarının Değerlendirilmesi*, Pamukkale Üniversitesi, Sosyal Bilimler Enstitüsü, Denizli.
- Chen, Y. S., Chen, L. H., & Takama, Y. (2015). Proposal of LDA-based sentiment visualization of hotel reviews. *In 2015 IEEE International Conference on Data Mining Workshop (ICDMW)* (pp. 687-693). IEEE.
- Ekinci, E., & Omurca, S. İ. (2017). Ürün özelliklerinin konu modelleme yöntemi ile çıkartılması. *Türkiye Bilişim Vakfı Bilgisayar Bilimleri ve Mühendisliği Dergisi*, 9(1), 51-58.
- Guo, Y., Barnes, S., J. & Jia, Q. (2017). Mining meaning from online ratings and reviews: tourist satisfaction analysis using latent dirichlet allocation. *Tourism Management*, 59(2017):467-483.
- Hu, N., Zhang, T., Gao, B., & Bose, I. (2019). What do hotel customers complain about? Text analysis using structural topic model. *Tourism Management*, 72, 417-426.
- Jadhav, N. (2014, 26 June). Topic models for sentiment analysis: A literature survey. *Center for Indian Language Technology (CFILT)*, 1-11.

- Maier, D., Waldherr, A., Miltner, P., Wiedemann, G., Niekler, A., Keinert, A., ... & Adam, S. (2018). Applying LDA topic modeling in communication research: Toward a valid and reliable methodology. *Communication Methods and Measures*, 12(2-3), 93-118.
- Nguyen, V. H., & Ho, T. (2021). Analyzing customer experience in hotel services using topic modeling. *Journal of Information Processing Systems*, 17(3), 586-598.
- Priyantina, R. A., & Sarno, R. (2019). Sentiment analysis of hotel reviews using latent dirichlet allocation, Semantic similarity and LSTM. *International Journal of Intelligent Systems*, 12(4), 142-155.
- www.rapidminer.com, Erişim Tarihi: 15.07.2022.
- www.tripadvisor.com, Erşim Tarihi: 15.07.2022.
- Xiang, Z., Du, Q., Ma, Y., & Fan, W. (2017). A comparative analysis of major online review platforms: Implications for social media analytics in hospitality and tourism. *Tourism Management*, 58, 51-65.

Analysis with Latent Dirichlet Allocation in the Evaluation of Hotel Services: The Case of Kastamonu

İbrahim BUDAK

Kastamonu University, Rectorate, Quality Coordinator, Kastamonu/Turkey

Alptekin SÖKMEN

Ankara Hacı Bayram Veli University, Faculty of Economics and Administrative Sciences, Ankara/Turkey

Extended Summary

Recently, a huge amount of content has been created on the web. Most of the customers post their opinions or ratings on review sites. From a managerial perspective, this data needs to be examined to understand what people think. However, analyzing such large amounts of data is time-consuming and cumbersome without filtering and classification methods. It is insufficient to display and describe all text data with a text-only interface. It is important to improve the performance of human cognition, to apply an effective classification and to represent the extracted information with visualization methods. One of the most important features for classifying reviews by users is tagging. LDA is a generative model for extracting tags based on hidden variables. Tags are represented as word bags that represent specific features of the corpus. LDA allocates word weights according to the probability distribution according to appropriate tags (Chen et al., 2015: 687). Today, customers can easily communicate with businesses through the website. Informal communication about the use and definition of certain products or services through internet technology is considered the most effective communication between service providers and consumers (Priyantina and Sarno, 2019: 142). In this study, customer comments on hotel services and products were analyzed with the LDA as a factor for hotel management evaluation and improvement, as well as a trigger for correcting errors and deficiencies in hotel products or services. Large and unstructured data were used to analyze the customer comments on the hotels' websites using appropriate techniques. Labels were created and visualized in line with the data.

The customer reviews of the hotels in Kastamonu on TripAdvisor are discussed. In the study, a total of 1687 customer evaluations were made until the second quarter of 2022. The open source RapidMiner 9.10.1 application was used to analyze the evaluation including customer comments, location, accommodation time, and travel type. Customer evaluations were analyzed with the help of data and text-processing operators in the application. It was seen that each user made an evaluation and the hotels were evaluated with an average overall score of 4.2.

Preprocessing: Customer comments have been converted into a digital form to be processed in the application. A numerical value was assigned to each different word in the comments and ranked according to their frequency. All operators and spokespersons in the application have been changed to lower-case characters in order to be in the same format. Words containing no meaning were filtered out to form a word bag. Words with less than three characters were filtered to reach the root of the word. Thus, the words to be labeled appropriately were reached.

Analysis: Customer comments on hotel services and products were analyzed with the LDA as a trigger factor to correct errors and omissions in the hotel's products or services, as well as being a factor for hotel management evaluation and improvement. Large and unstructured data were used to analyze the customer reviews written on the

websites of the hotels using appropriate techniques. In the next stage, 5 labels were created and visualized in line with the data. These tags are: essential services (0.271), experience (0.212), comfort and convenience (0.183), value and fun (0.169), things to do (0.164). Here are the 10 most frequently used words for recent tags.

Collecting and improving numerical and textual data in the digital environment is both faster and less costly. For this reason, when compared to the data collected by research companies by different methods, the volume of data that users share in the digital environment in line with their own wishes is large and objective. It is thought that the analysis of the data obtained about the hotels in Kastamonu will give important information about achieving strategic goals in a competitive environment. With the tags and frequently used words revealed by the analysis, risks and opportunities will be evaluated. In future studies, the situations of hotels located in different provinces can be compared and analyzed in a wider perspective.