



# Veri Madenciliği Teknikleri Yardımıyla Otel Yorumlarından Anahtar Kelimeler Keşfi

Arif KOYUN<sup>\*1</sup>, Didem YANGEÇ<sup>2</sup>

<sup>1</sup> Süleyman Demirel Üniversitesi, Mühendislik Fakültesi, Bilgisayar Mühendisliği Bölümü-Isparta, Türkiye

<sup>2</sup> Süleyman Demirel Üniversitesi, Fen Bilimleri Enstitüsü, Bilgisayar Mühendisliği Bölümü Yüksek Lisans öğrencisi Isparta, Türkiye

(İlk Geliş Tarihi 7 Kasım 2018 ve Kabul Tarihi 28 Kasım 2018)

(DOI: 10.31590/ejosat.479833)

## Öz

Günümüzde teknolojinin hızlı gelişimi ile bilinçlenen ve sorgulamaktan kaçınmayan tüketiciler, şikâyetlerini işletmelere iletme süreci içine girmektedirler. Bu çalışmada İzmir ilinin sınırları içerisinde yer alan 92 adet otele ilişkin şikâyetlerin veya tavsiyelerin yapıldığı web sitesinden otellere yapılan yorumları çekerek MySQL veri tabanına kaydedilmesini sağlayan PHP bot yazılımı gerçekleştirilmiştir. Matlab ortamında geliştirilen veri madenciliği yöntemlerinden biri olan veri görselleştirme yöntemi kullanılarak verinin analizi gerçekleştirilmiştir. Spring Tool Suite ortamında aranan anahtar kelime en çok hangi otelde kullanılmış, her otelin veri analiz (grafik) sonuçlarının gösterildiği web uygulaması oluşturulmuştur. Geliştirilen bu uygulamayla müşteriler her otele yapılan sayısız yorumları tek tek okumak zorunda kalmadan otelin öne çıkan özelliklerini öğrenebileceklerdir. Müşteriler hem zaman tasarrufu sağlamış hem de otel hakkında önceden fikir sahibi olmuş olacaktır. İşletme sahipleri de müşteri taleplerini göze alarak hizmetlerini geliştirebilecektir.

**Anahtar Kelimeler:** Otel tercihi, Veri madenciliği, Seyahat davranışı, Gelişmekte olan model madenciliği

## Discovery of Keywords from Hotel Comments with the Help of Data Mining Techniques

### Abstract

Today consumers, who are conscious of the rapid development of technology and who do not abstain from questioning, are in the process of communicating their complaints to institutions. In this study, PHP bot software was implemented which enables to record into MYSQL database by pulling comments from the web site where complaints and recommendations were made about the 92 hotels locating within the borders of İzmir. Data analysis was performed using data visualization method which is one of the data mining methods developed in Matlab stage. In the Spring Tool Suite environment, a web site was created that shows the data analysis (graphic) results of each hotel to find out which hotel was mostly used. With this application developed, customers will be able to read the numerous comments of each hotel individually and learn about the highlights of the hotel. The customers both have saved time and have already had an idea about the hotel. Business owners will be able to improve their services by receiving customer requests.

**Keywords:** Hotel preference, Data mining, Travel behavior, Emerging pattern mining

### 1. Giriş

Veri madenciliği teknikleri kullanılarak bir hizmet sektörü olan konaklama tesislerine yapılan yorumları ele alarak veri ya da metin analizi gerçekleştiren çok sayıda araştırma vardır. Lau ve ark., 2005, Hong Kong'daki otellerden veriler toplanılarak metin madenciliği analizi gerçekleştirilmiştir. Araştırma için üç çalışma gerçekleştirilmiştir. Birincisi otelin profilini (hizmetler, yer, iletişim) içeren bir veri tabanı oluşturulmuştur, İkinci çalışma otellerin fiyat çizelgesinin veri tabanı oluşturulmuştur. Üçüncü çalışma ise internet üzerinde otellere yapılan yorumlar incelenmiştir. Araştırma sonucunda otelciler tarafından rekabet ortamı ve strateji geliştirmiştir.

Tian ve ark., 2016, Çin'in dört büyük şehrindeki 58 adet 3 ile 5 yıldızlı otelin TripAdvisor sitesindeki İngilizce yorumlarını analiz etmişlerdir. Geliştirilen bir program ile yorumlardaki anahtar kelimeleri ayrıştırıp excel tablosu haline getirilmiştir. Daha sonra bu verileri farklı grafikler oluşturmak suretiyle analiz etmişlerdir. Bu sayede, hangi konularda müşterilerin olumlu ve olumsuz düşündükleri hakkında hotel yönetimlerine bilgi verilmesi amaçlanmıştır.

Güzel, 2014, araştırmada tripadvisor.com sitesinde marka prestijini korumak için puanlama açısından ilk sırada yer alan örnek bir konaklama işletmesi ele alınarak, otele yapılan şikâyetler değerlendirilmiş ve nitel araştırma yöntemlerinden biri olan içerik analizi gerçekleştirilmiştir. İçerik analizi sonucunda

hizmetten memnun kalmayan müşterilerin konaklama tesisine yönelik kötüleme ifadelerini kullandıkları, tavsiye etmedikleri ve başka otele yönlendirme ifadeleri kullandıkları da keşfedilmiştir.

Han ve ark., 2016 tarafından yapılan araştırmalarda olumsuz yorumların bir otelin konuk değerlendirmesinde olumlu yorumlara göre daha büyük bir ağırlığa sahip olduğunu tespit etmişlerdir. Bu yanlış değerlendirme, basit bir pozitif ve negatif puanın ortalaması, konukların otel hakkındaki görüşlerini net bir şekilde göremeyebileceği anlamına gelmiştir. Çalışmada, etkinlik, deneyim, yer, işlem ve değer ile ilgili beş özelliğin 18.106 ayrı terimin ilişkileri analiz edilmiştir. Farklı otel katmanları için yapılan incelemeler, bu niteliklere göre değişen ağırlıklar vermiştir. Alt katmanlı mülklerdeki konuklar, otellerde orta ve üst katlarda kalanlardan daha fazla işlem ve değer hakkında yazmışlardır. İncelemelerin özellikle göze çarpan bir özelliği, kısa yorumlarda yüksek değerlendirme yapılırken, uzun yorumlarda değerlendirme puanının azaldığı fark edilmiştir. Konukların memnuniyetsizlik hissettiklerinde değer ve işlemlerle ilgili daha fazla bilgi yazdıklarını da bulmuşlardır. Bu nedenle, metin analizi, yöneticilerin misafirlerin değerlendirmelerini ve otellerin derecelendirmelerini iyileştirmek için alabilecekleri belirli adımları gösterebileceğini tespit etmişlerdir. Yetkin ve ark., 2016, Alanya bölgesinde hizmet sunan 48 konaklama tesisinde çalışan 223 otel yöneticilerinin oteli değerlendirme sitelerine yönelik otel yöneticilerinin tutumlarını araştırmıştır. Konaklama tesislerini değerlendirme sitelerinin en önemli kaynaklarından biri ağızdan ağıza iletişimidir. Bu durum konaklama tesislerini değerlendirme siteleri, turistlerin tercihlerini etkilemektedir. Araştırma sürecinden elde edilen sonuçlardan otel yöneticilerinin konaklama tesislerini değerlendirme sitelerinin varlığından haberdar olduğu anlaşılmıştır. İşletme sahibi, genel müdür, müşteri ilişkileri birimi ile ön büro birimi yöneticilerinin konaklama tesislerini değerlendirme sitelerini düzenli olarak takip eden yöneticiler olduğu belirlenmiştir. Yiyecek ve içecek birimi, muhasebe ve kat hizmetleri yöneticilerinin diğer birim yöneticilerine göre konaklama tesislerini değerlendirme sitelerinde yer alan tavsiye ve şikâyetleri daha az okuduklarını saptamıştır.

Hilage ve Kulkarni, 2012 ise veri madenciliği, veri tabanlarından veri madenciliği yapmak ve veri tabanından anlamlı desenler bulmak için kullanmışlardır. Birçok organizasyon şu anda bu veri madenciliği tekniklerini kullanmaktadır. Yazarlar, ortaklık kuralları, kural indüksiyon tekniği, apriori algoritması, karar ağacı ve sinir ağı gibi veri madenciliği tekniklerinin literatürünü gözden geçirmişlerdir. Yapılan inceleme, veri madenciliği tekniklerinin, veri tabanından anlamlı bir model bulmak için farklı uygulama alanları için nasıl kullanıldığı üzerine odaklanmıştır.

Teknolojinin gün geçtikçe gelişmesi, teknolojik aletlerin (bilgisayar, tablet vb.) ucuzlaması ve giderek artması, internet erişiminin artması, veri depolama sistemlerinin yaygınlaşması çok sayıda verinin kaydedilerek saklanmasını ucuz ve kolay hale getirmiştir. Bu çok sayıdaki veri işlenmemiş, özümlememiş gerçekler olarak tanımlanmaktadır. Değersiz gibi görünen verinin anlam kazanabilmesi için bir amaç doğrultusunda işlenmesi gerekmektedir. Her geçen gün artan bu verilerin analizinde geleneksel teknikler yetersiz kalmıştır. Verilerin anlamlı hâle dönüştürülmesi için yeni teknik ve araçlara ihtiyaç duyulmuştur. Bu ihtiyaçlara bağlı olarak veri tabanı ve bilgisayar teknolojilerini bir araya getiren veri madenciliği ortaya çıkmıştır. Özbay, 2015, Can ve ark., 2012.

## 2. Materyal ve Metot

Veri madenciliği en basit tanımıyla, çok büyük miktardaki veri yığınlarının amaca uygun olarak anlamlı hale dönüştürülmesidir. Literatürlerde yer alan veri madenciliğini şu şekilde tanımlayabiliriz: veri madenciliği, veri yığınlarının belli bir amaç doğrultusunda işlenerek anlamlı bilginin elde edilmesidir Aydılek ve ark., 2015. Veri madenciliği, depolarda depolanan büyük miktarda veriyi eleyerek, kalıp tanıma teknolojilerini ve istatistiksel ve matematiksel teknikleri kullanarak anlamlı yeni bağıntıları, desenleri ve eğilimleri keşfetme sürecidir Sharma, 2014. Veri madenciliği, enformasyon keşfi, veri arkeolojisi, bilgi aktarımı olarak adlandırılarak, saklı kalmış değerli ve kullanılabilir veriyi ortaya çıkarma sürecidir Koyuncuğil ve Özgülbaş, 2009. Bu tanımlardan yola çıkarak genel tanım yapmak mümkündür: Veri madenciliği, karmaşık ve büyük veriyi, yapay zekâ, istatistik, veri tabanlarında bilgi keşfi, makine öğrenmesi, bilgisayar bilimi gibi disiplinlerden faydalanarak amaca uygun modellerin ortaya çıkarılması işlemidir.

Bu çalışmada tüketicilerin şikâyetlerinin veya tavsiyelerinin yapıldığı web sitesinden otellere yapılan yorumların MySQL veri tabanına kaydedilmesini sağlayan bir yazılım geliştirilmiştir. Bu amaçla PHP bot yazılmıştır. İnternette bir amaç için kodlar aracılığı ile istenen adrese ulaşip oradan istenilen içeriği çekmeye, saklamaya, içeriği alıp istenen yerlerde yayınlamaya yarayan robotlara bot denilmektedir. Botlar bilgiyi, veri çekmek istediğimiz adresten o adreste bulunan birçok etiket aracılığıyla çekmektedir.

Veri madenciliği yöntemlerinden biri olan veri görselleştirme yöntemi kullanılarak verilerin analizi gerçekleştirilmiştir. Veri görselleştirme, verinin hızlı ve kolay bir şekilde keşfedildiği ve veri hakkında genel bir kaniya varıldığı için oldukça kullanışlı bir yöntemdir. Veri görselleştirme de veriler grafikler ile sunulmaktadır. Matlab ortamının sağladığı ileri görselleştirme ve algoritma geliştirme özelliği nedeniyle bu çalışmada Matlab kullanılmıştır. Matlab ortamında geliştirilen yazılımla, yapılan yorumların analizi gerçekleştirilmiştir.

Geliştirme ortamı olarak Spring Tool Suite, programlama dili olarak Java kullanılarak, kullanıcılara otele ilişkin anahtar kelimelerin istatistiklerini, veri analiz sonuçlarını sunan bir web uygulaması geliştirilmiştir. Spring Java uygulamalarını düzgün bir alt yapıyla hızlı bir şekilde geliştirmek için kullanılan açık kaynaklı modüler bir uygulama çatısıdır. Web uygulamaları geliştirmek için Java EE (Enterprise Edition) üzerine inşa edilmiş MVC (Model View Controller) Projesi ve AOP (Aspect Oriented Programming) Modülü vardır.

### 2.1. Veri Ambarının Oluşturulması

Bu çalışmada, veri madenciliği sürecinde kullanılacak olan ham veri, bir hizmet sektörü olan konaklama tesislerini değerlendirme sitesinden elde edilmiştir. Öncelikle PHP bot uygulaması geliştirilmiş, verinin MySQL veri tabanına kaydı gerçekleştirilmiştir.

PHP dilinde bot yazarken iki yöntem vardır:

- 1) File\_get\_contents
- 2) Curl

Bu çalışmada, daha güvenli ve hızlı olması avantajıyla curl yöntemi kullanılmıştır.

## Preg\_match ve Preg\_match\_all Fonksiyonu

Bu çalışmada yer alan preg\_match fonksiyonunun bir bölümü Şekil 1'de gösterilmiştir.

```
preg_match('~<span class="fn" id="hp_hotel_name">(.*?)</span>~s', $output, $otelismix);
preg_match_all('~<li class="clearfix one_col" data-question="(?:.*?)" >(.*?)</li>~s', $degerlendirmeler[1], $degerlendirmelistesi);
```

Şekil 1. Preg\_match ve preg\_match\_all fonksiyonları

Preg\_match ve preg\_match\_all fonksiyonları belirtilen ifadeyle içeriğin eşleşmesini kontrol eder. İkisi arasındaki fark preg\_match eşleşen tek kaydı getirir. Preg\_match\_all ise eşleşen bütün kayıtları getirir.

## PHP Düzenli İfadeler (Regular Expression/Regex) Kullanımı

Düzenli ifadeler, eşleştirme yapmak için kullanılan programdır. Birçok programlama dilinde kullanılan düzenli ifadeler en çok PHP bot yapımında kullanılır.

## 2.2. Veri Madenciliği Çalışmaları

Çalışmanın bu aşamasında verinin toplanması, veri ön işleme ve özellik çıkarımı gerçekleştirilmiştir.

İzmir ilinin sınırları içerisinde yer alan 92 otel tesise yapılan toplam 25564 adet müşteri yorumu ele alınmıştır. Bu yorumlar müşterilerin konaklama süresi içerisindeki memnuniyetleri ve şikâyetleri içermektedir. Bunun için Bölüm 2'de anlatılmış olan PHP bot uygulaması geliştirilerek bu yorumların veri tabanına kaydı gerçekleştirilmiştir. Çalışma kapsamında MySQL veri tabanına kaydedilen veri üzerinde analiz gerçekleştirilmiştir.

Çalışma için elde edilen veri, bilgisayar tarafından işlenmeyen ham veridir. Ham veri üzerinde analiz ve çıkarımlar yapabilmek için bu verinin bazı ön işlemlerden geçmesi gerekmektedir. Sağladığı ileri görselleştirme ve algoritma geliştirme özelliğine bağlı olarak, çalışmanın veri ön işleme, analiz ve histogram bilgisinin çıkarımı yapılmıştır. Müşteri yorumları üzerinde şu işlemler uygulanmıştır:

- küçük harfe dönüştürme
- boşluk silme
- noktalama işaretlerini silme
- edatları ve bağlaçları ayıklama (ancak, ve, fakat, için, and, the vb.)

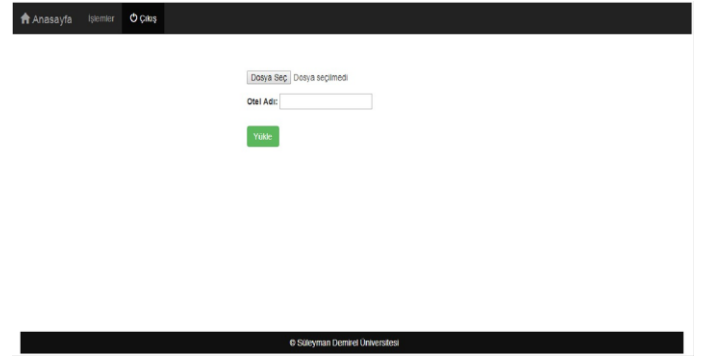
Çalışmamızda bu adıma kadar veri kümesini belirleyip verileri topladık, veriyi ön işleme tabi tutarak temizledik. Daha sonra her otel için şu adımlar gerçekleştirilmiştir:

- yapılan yorumlarda geçen kelimelerin ve sıklıklarının unique fonksiyonu ile belirlenmesi
- bu kelimelerin tekrar sayılarına göre azalan şekilde sort fonksiyonu ile sıralanması
- kelime sıklıklarının yüzdelerinin hesaplanması
- elde edilen sonuçların (geçen her anahtar kelime için kelime, tekrar sayısı ve yüzdesi) tablo haline getirilmesi
- en çok geçen anahtar kelimelerin histogram olarak görselleştirilmesi

## 2.3. Geliştirilen Web Uygulaması

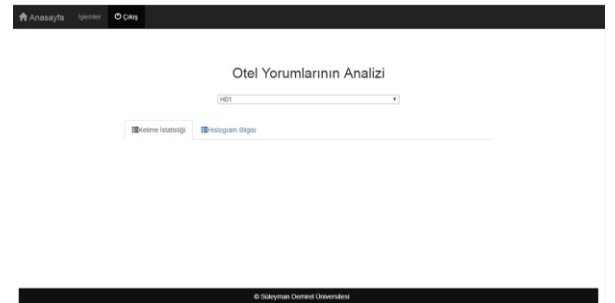
Çalışmanın bu aşamasında elde edilen sonuçların, geliştirilen web uygulanmasında sunulması gerçekleştirilmiştir. Veri madenciliği çalışmalarıyla elde edilen anahtar kelimelerin sıklığı, veri görselleştirme yöntemiyle elde edilen histogram bilgilerinin, sunulması için Spring Tool Suit (STS) geliştirme ortamında Java programlama diliyle geliştirilen bir web uygulaması oluşturulmuştur. Oluşturulan ilk arayüz, her otelin kelime sıklık ve histogram bilgilerinin kaydı için geliştirilmiştir. Kayıt işlemleri için veri tabanı tasarlanmıştır. Oluşturulan ikinci arayüz, müşterilerin gitmek istediği otelin daha önce hangi niteliğiyle ön planda olduğu hakkında bilgi sahibi olacağı bir arayüz geliştirilmiştir.

Otel adı ve analiz sonuçlarının girildiği web arayüzü Şekil 2'de verilmiştir. Girilen bu bilgiler veri tabanına kaydedilmiştir.

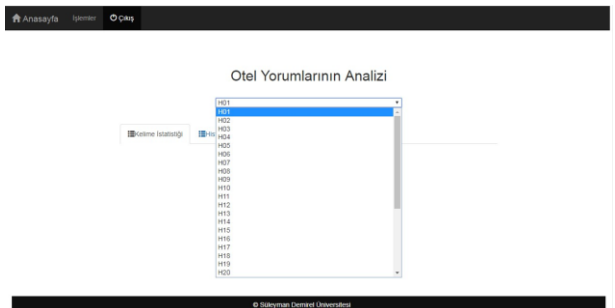


Şekil 2. Otel bilgisinin girildiği web arayüzü

Seçilen otele ilişkin yorumların analizi (kelime istatistiği ve histogram bilgisi) ile ilgili arayüz ise Şekil 3 ve 4'de görülmektedir.



Şekil 3. Otel yorumlarının analiz sonuçlarını gösteren web arayüzü

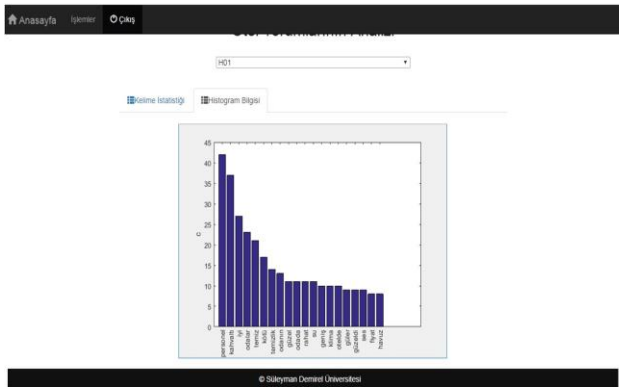


Şekil 4. Listedten otel seçimi

Seçilen otele ilişkin kelime istatistiği Şekil 5’de, histogram bilgisi Şekil 6’da görülmektedir.

AnahtarKelimeler	TekrarSayisi	KelimeYuzdesi
'personel'	42	1.5843
'kahvaltı'	37	1.3957
'iyi'	27	1.0305
'odalar'	23	0.8676
'temiz'	21	0.78215
'konu'	17	0.64127
'temizlik'	14	0.5291
'odaları'	13	0.49028
'güzel'	11	0.41494
'odada'	11	0.41494
'rahat'	11	0.41494
'su'	11	0.41494
'geniş'	10	0.37722
'kilime'	10	0.37722
'otelde'	10	0.37722
'güler'	9	0.33949
'güzelidi'	8	0.30176
'işer'	8	0.30176

Şekil 5. Otel kelime istatistiği



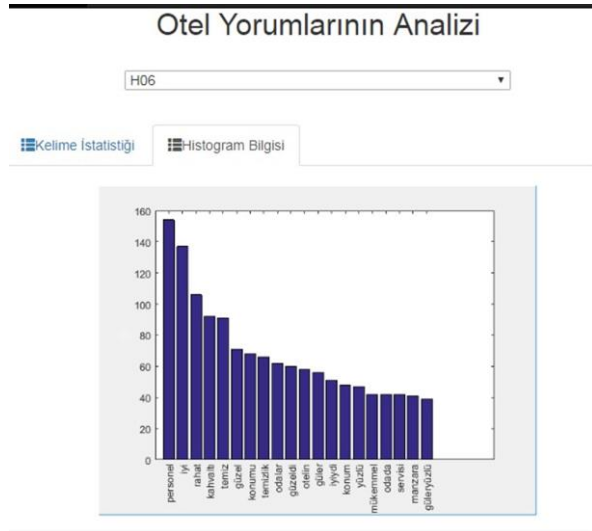
Şekil 6. Otel histogram istatistiği

### 3. Araştırma Bulguları ve Tartışma

Çalışmanın bu aşamasında geliştirilen uygulamalar neticesinde elde edilen sonuçlar analiz edilip yorumlanmıştır. Çalışmada İzmir ilinin sınırları içerisinde yer alan 92 otele ait müşteri yorumlarından veri seti elde edilerek, otellere ilişkin anahtar kelimelerin keşfi gerçekleştirilmiştir. Bununla birlikte otellere ilişkin anahtar kelimelerin istatistikleri ve veri analiz sonuçlarını sunan bir web uygulaması geliştirilmiştir. Bu uygulamada müşteriler tarafından daha çok tercih edilen ve analiz sonuçlarını daha anlamlı yorumlanmasını sağlayan H06, H07, H09, H18, H21, H22, H28, otellerine ait kelime istatistikleri ve histogram bilgileri sırasıyla Şekil 7, Şekil 8, Şekil 9, Şekil 10, Şekil 11, Şekil 12, Şekil 13, Şekil 13, Şekil 15, Şekil 16, Şekil 17, Şekil 18, Şekil 18, Şekil 20’de gösterilmiş ve yorumlanmıştır.

AnahtarKelimeler	TekrarSayisi	KelimeYuzdesi
'personel'	154	1.2211
'iyi'	137	1.0863
'rahat'	106	0.84047
'kahvaltı'	92	0.72946
'temiz'	91	0.72154
'güzel'	71	0.56296
'konumu'	68	0.53917
'temizlik'	66	0.52331
'odalar'	62	0.4916
'güzelidi'	60	0.47574
'otelin'	58	0.45988
'güler'	56	0.44402
'iyiydi'	51	0.40438
'konum'	48	0.38059
'yüzlü'	47	0.37266
'mükemmel'	42	0.33302
'odada'	42	0.33302
'servisi'	42	0.33302
'manzara'	41	0.32509
'güleriylü'	39	0.30923

Şekil 7. H06 oteline ait kelime istatistiği



Şekil 8. H06 oteline ait histogram bilgisi

H06 otelini tercih eden müşterilerin personellerin ilgisinden memnun kaldıkları, temizlik hizmetinin, konumunun ve manzarasının iyi olduğu sonucuna varabiliriz.

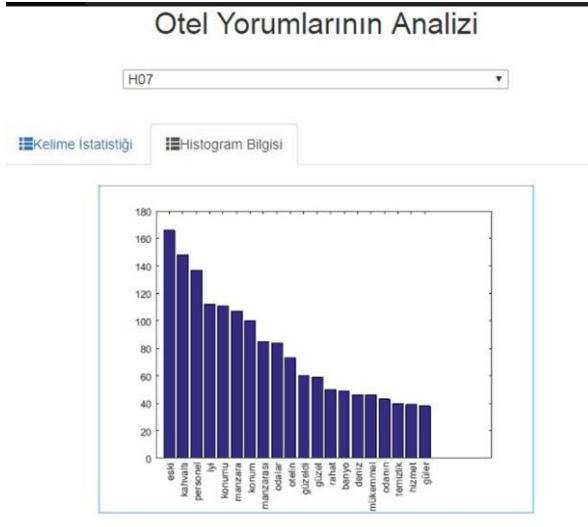
AnahtarKelimeler	TekrarSayisi	KelimeYuzdesi
'eski'	166	1.2559
'kahvaltı'	148	1.1197
'personel'	137	1.0365
'iyi'	112	0.84733
'konumu'	111	0.83976
'manzara'	107	0.8095
'konum'	100	0.75654
'manzarası'	85	0.64306
'odalar'	84	0.6355
'otelin'	73	0.55228
'güzeldi'	60	0.45393
'güzel'	59	0.44636
'rahat'	50	0.37827
'banyo'	49	0.37071
'deniz'	46	0.34801
'mükemmel'	46	0.34801
'odanın'	43	0.32531
'temizlik'	40	0.30262
'hizmet'	39	0.29505
'gülür'	38	0.28749

Şekil 9. H07 oteline ait kelime istatistiği

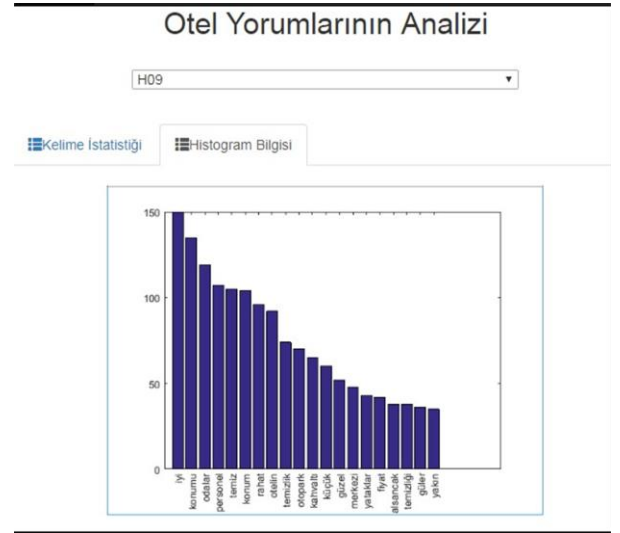
Otele ait anahtar kelimeleri inceleyecek olursak, otel yöneticilerinin ve personellerin ilgisinden memnun kaldığı, kahvaltı, temizlik hizmetinin iyi olduğu, manzarası ve konumunun güzel olduğu sonucunu çıkarabiliriz. Eski kelimesinin tekrar sayısının fazla olmasına bağlı olarak bu otelin eski olduğu sonucuna varılabilir.

AnahtarKelimeler	TekrarSayisi	KelimeYuzdesi
'iyi'	150	1.3256
'konumu'	135	1.193
'odalar'	119	1.0516
'personel'	107	0.94556
'temiz'	105	0.92789
'konum'	104	0.91905
'rahat'	96	0.84836
'otelin'	92	0.81301
'temizlik'	74	0.65394
'otopark'	70	0.61859
'kahvaltı'	65	0.57441
'küçük'	60	0.53022
'güzel'	52	0.45953
'merkezi'	48	0.42418
'yataklar'	43	0.37999
'fiyat'	42	0.37116
'alsancak'	38	0.33581
'temizliği'	38	0.33581
'gülür'	36	0.31813
'yakın'	35	0.3093

Şekil 11. H09 oteline ait kelime istatistiği



Şekil 10. H07 oteline ait histogram bilgisi



Şekil 12. H09 oteline ait histogram bilgisi

Fiyat kelimesinin tekrar sıklığının fazla olması H09 oteline tercih eden müşterilerin fiyattan memnun kaldıklarını göstermektedir. Çünkü otelde iyi kelimesinin tekrar sıklığı fazladır. Otele yapılan yorumlarda iyi veya kötü kelimeleri yapılan yorumlarla paraleldir. Merkez, konum ve yakın kelimelerinden bu otelin merkeze yakın olduğu, otopark sıkıntısının olmadığı, otelin temiz olduğu sonucuna varabiliriz.



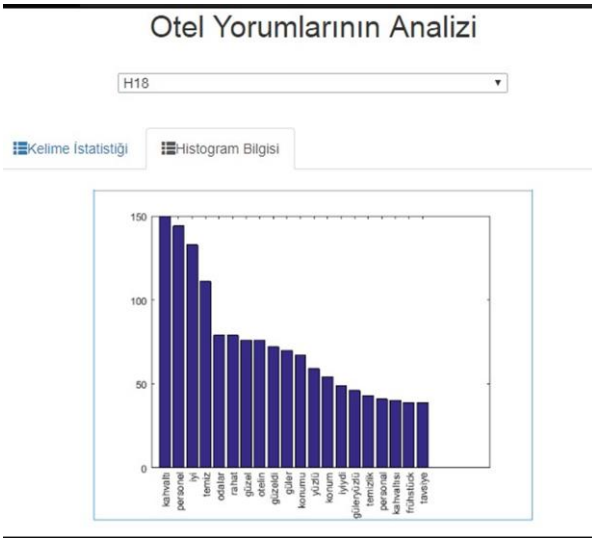
H18

Kelime İstatistiği Histogram Bilgisi

AnahtarKelimeler	TekrarSayisi	KelimeYuzdesi
'kahvaltı'	150	1.0543
'personel'	144	1.0121
'iyi'	133	0.93478
'temiz'	111	0.78015
'odalar'	79	0.55524
'rahat'	79	0.55524
'güzel'	76	0.53416
'otelin'	76	0.53416
'güzeldi'	72	0.50604
'güler'	70	0.49199
'konumu'	67	0.4709
'yüzlü'	59	0.41468
'konum'	54	0.37953
'iyiydi'	49	0.34439
'gülyüzlü'	46	0.32331
'temizlik'	43	0.30222
'personal'	41	0.28816
'kahvaltısı'	40	0.28114
'frühstück'	39	0.27411
'tavsiye'	39	0.27411

Şekil 13. H18 oteline ait kelime istatistiği

Otele ait anahtar kelimeleri inceleyecek olursak, iyi kelimesinin sıklığının fazla olması nedeniyle otelin kahvaltı, temizlik hizmetlerinin iyi olduğu, personelin gülyüzlü ve ilgili olduğunu, odaların rahat, otel konumunun iyi olduğunu sonucunu çıkarabiliriz.



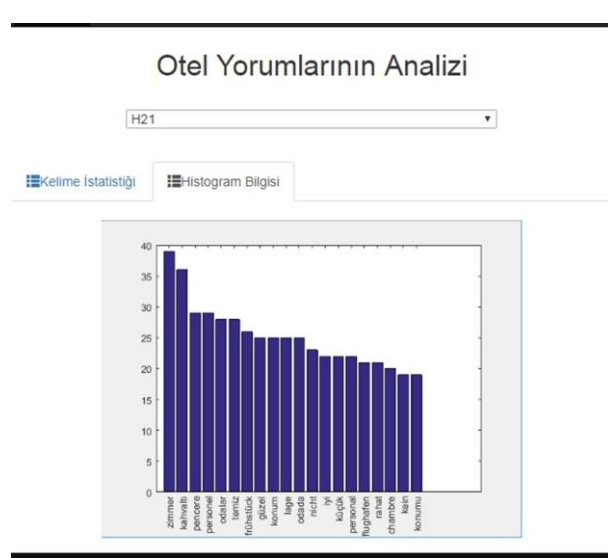
Şekil 14. H18 oteline ait histogram bilgisi

H21

Kelime İstatistiği Histogram Bilgisi

AnahtarKelimeler	TekrarSayisi	KelimeYuzdesi
'zimmer'	39	0.49063
'kahvaltı'	36	0.45289
'pencere'	29	0.36483
'personel'	29	0.36483
'odalar'	28	0.35225
'temiz'	28	0.35225
'frühstück'	26	0.32709
'güzel'	25	0.3145
'konum'	25	0.3145
'lage'	25	0.3145
'odada'	25	0.3145
'nicht'	23	0.28934
'iyi'	22	0.27676
'küçük'	22	0.27676
'personal'	22	0.27676
'flughafen'	21	0.26418
'rahat'	21	0.26418
'chambre'	20	0.2516
'kein'	19	0.23902
'konumu'	19	0.23902

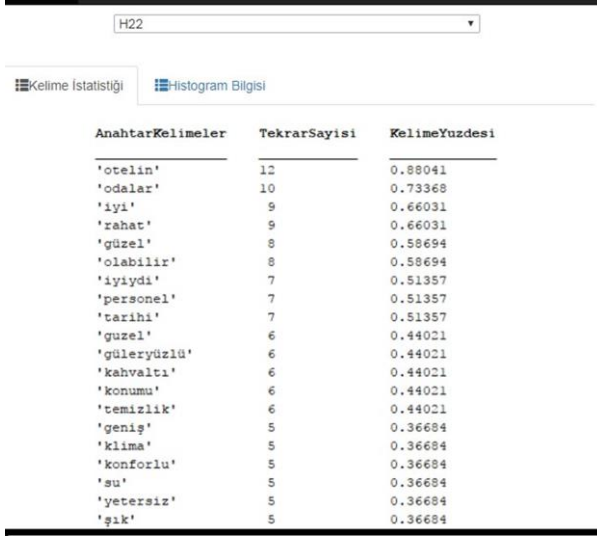
Şekil 15. H21 oteline ait kelime istatistiği



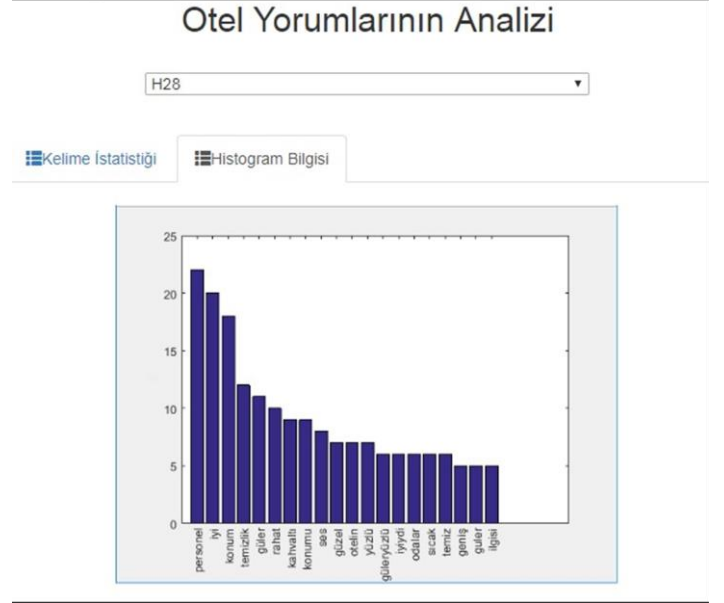
Şekil 16. H21 oteline ait histogram bilgisi

Zimmer, frühstück, flughafen anahtar kelimeleri Almanca kelimeler olup Türkçe karşılıkları sırasıyla oda, kahvaltı, havaalanıdır. Chambre anahtar kelimesi Fransızca olup Türkçe karşılığı odadır. H21 oteli, yabancı uyruklu vatandaşların daha çok tercih ettiği bir otel olup, en çok oda ve kahvaltı hizmeti hakkında yorum yapılmıştır. Odaların geniş ve rahat olduğu, müşterilerin hizmetten memnun kaldıkları sonucuna varabiliriz.

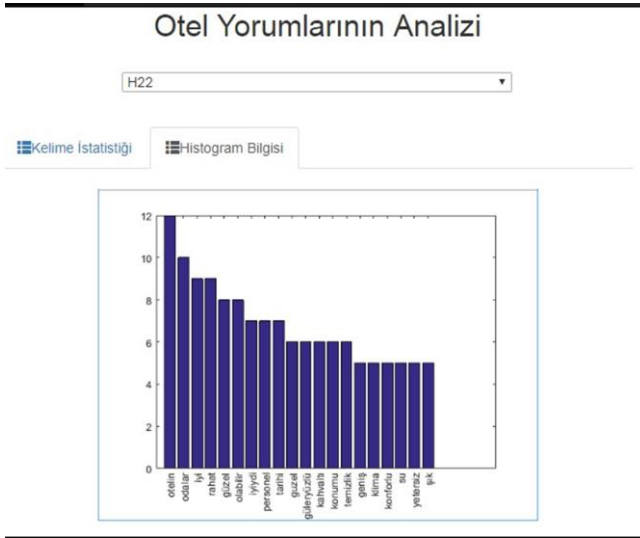
H22 oteli için Şekil 17'de görüldüğü gibi anahtar kelimelerin sıklığı azdır. Bu otelin genel olarak iyi bir otel olduğu, fakat konaklamak için pek tercih edilen bir otel olmadığı gözlemlenmiştir. H22 otelinin histogram bilgisi Şekil 18'de gösterilmiştir.



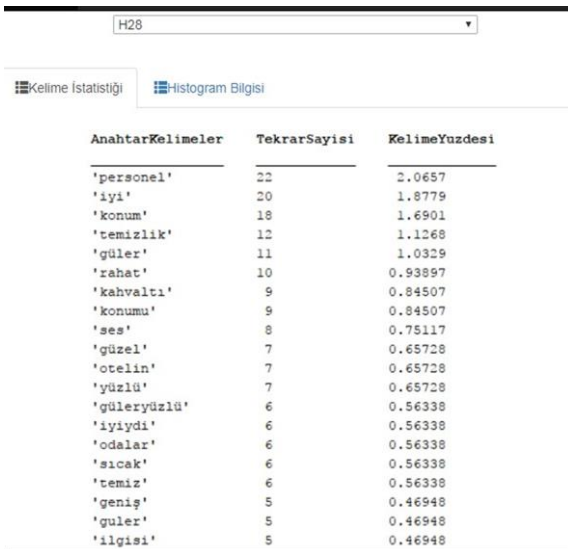
Şekil 17. H22 oteline ait kelime istatistiği



Şekil 20. H28 oteline ait histogram bilgisi



Şekil 18. H22 oteline ait histogram bilgisi



Şekil 19. H28 oteline ait kelime istatistiği

H28 otelinin kelime istatistik bilgisini incelersek, personelin güleryüzlü ve ilgili olduğu, odaların temiz ve geniş, yatakların rahat, otel konumunun iyi olduğu sonucuna varabiliriz.

#### 4. Sonuç

Teknoloji araçları ve internet hızlı bir gelişim göstermiştir. Bugün çoğu işlemin internet üzerinden yapılabilir olması kullanıcı sayısını artırmıştır. Bir hizmet sektörü olan turizm sektöründe sürekli gelişen internet önemli bir iletişim aracı olmuştur. Konaklama tesislerini değerlendirme siteleri, müşteriler tarafından konaklama öncesi ve sonrası memnuniyet, şikâyet ve tavsiyelerini aktarma aracı olarak kullanılmaktadır. Bu nedenle konaklama tesislerini değerlendirme sitelerinde sayısız veri yığınları ve ciddi oranda büyük veri oluşmaktadır.

Bu çalışmada, müşterilerin tavsiyelerinin veya şikâyetlerinin yapıldığı web sitesinden otellere yapılan yorumlar, geliştirilen PHP bot uygulaması aracılığıyla MySQL veri tabanına kaydedilmiştir. Matlab'ın veri tabanı explorer uygulaması aracılığıyla MySQL JDBC bağlantısı ayarlanıp veri tabanına bağlanılır. Toplanan yorum metinleri, küçük harfe dönüştürme, boşluk silme, notalama işaretlerini boşluğa çevirme ve edat ve bağlaçları ayıklama gibi bazı ön işlemlerden geçirilmiştir. Yorum metinleri, boşluklardan bölünerek kelimelere ayrıştırılmıştır. Yapılan yorumlarda en çok tekrar eden anahtar kelimelerin sıklıkları ve bu kelimelerin yüzdeleri hesaplanmıştır. Ayrıca en çok tekrar eden benzersiz kelimeler histogram olarak görüntülenmiştir.

Matlab ortamında elde edilen sonuçların veri tabanına kaydedilmesi ve kullanıcıların otellere ilişkin yorumlarda yer alan anahtar kelimeleri histogram olarak görüntüleyebilmeleri için Java tabanlı bir web uygulaması geliştirilmiştir. Geliştirilen bu uygulama aracılığıyla müşteriler, otellere ilişkin istatistik bilgileri ve analiz sonuçlarını kolaylıkla görebilmektedirler. Müşteriler tarafından tercih edilen otellere yapılan yorum sayısı fazla olduğu ve yorum sayısı az olan otellerin müşteriler tarafından talep görmediği anlaşılmıştır. Otellere yapılan yorum sayısı müşterinin tercihini etkilediği görülmüştür.

## **Kaynaklar**

- Aydilek, İ.B., Ayneli, D., Uçan, M., 2015. Veri Madenciliği İle Bilgisayar Mühendisliği Öğrencilerinin Bölüme Olan Yatınlığının Değerlendirilmesi, Akademik Bilişim-2015, Eskişehir.
- Can, M.B., Çamur, E., Koru, M., Özkan, Ö., Rzayeva, Z., 2012. Veri Kümelerinden Bilgi Keşfi: Veri Madenciliği, Başkent Üniversitesi Tıp Fakültesi XIV. Öğrenci Sempozyumu, Ankara.
- Güzel, F.Ö., 2014. Marka İtibarını Korumada Şikâyet Takibi: Çevrimiçi Seyahat 2.0 Bilgi Kanallarında Bir Uygulama, *Journal of Internet Application and Management* 5(1), 5-19.
- Han, H.J.S., Mankad S., Gavirneni N., Verma R., 2016. What Guests Really Think of Your Hotel: Text Analytics of Online Customer Reviews, *Cornell Hospitality Report* 16(2), 3-17.
- Hilage T. A., Kulkarni R. V., 2012. Review of Literatüre on Data Mining, *International Journal of Research and Reviews in Applied Sciences* 10(1), 107-114.
- Lau K., Lee K., Ho Y., 2005. Text Mining for the Hotel Industry, *Cornell University* 46(3), 344-362.
- Koyuncugil, A.S, Özgülbaş, N., 2009, Veri Madenciliği: Tıp ve Sağlık Hizmetlerinde Kullanımı ve Uygulamaları, *Bilişim Teknolojileri Dergisi* 2(2), 21-32.
- Sharma, M., 2014. Data Mining: A Literature Survey, *International Journal of Emerging Research in Management and Technology* 3(2), 1-4.
- Özbay, Ö., 2015. Veri Madenciliği Kavramı ve Eğitimde Veri Madenciliği Uygulamaları, *Uluslararası Eğitim Bilimleri Dergisi* 5, 262-272.
- Yetkin M., Üngüren E., Kaçmaz Y.Y., 2016. Otel Yöneticilerinin Otel Değerlendirme Sitelerine Yönelik Tutumları Üzerine Bir Araştırma, *Uluslararası Yönetim İktisat ve İşletme Dergisi* 12(28), 99-122.
- Tian, X., He, W., Tao, R., & Akula, V., 2016. Mining Online Hotel Reviews: A Case Study from Hotels in China. Twenty-second Americas Conference on Information Systems, San Diego-USA.