



**KTO KARATAY ÜNİVERSİTESİ  
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ  
ELEKTRİK BİLGİSAYAR MÜHENDİSLİĞİ ANABİLİM DALI  
TEZLİ YÜKSEK LİSANS PROGRAMI**

**DOĞAL DİL İŞLEME VE KÜMELEME YÖNTEMLERİ İLE MÜŞTERİ  
YORUMLARININ İNCELENMESİ**

**Recep Ali AY**

**Yüksek Lisans**

**KONYA  
Mart 2021**

DOĐAL DİL İŐLEME VE KÜMELEME YÖNTEMLERİ İLE MÜŐTERİ  
YORUMLARININ İNCELENMESİ

Recep Ali AY

KTO Karatay Üniversitesi  
Lisansüstü Eğitim Enstitüsü  
Elektrik Bilgisayar Mühendisliđi Anabilim Dalı  
Yüksek Lisans Programı

Yüksek Lisans

Tez Danışmanı: Dr. Öğr. Üyesi Semih YUMUŐAK

Konya  
Mart 2021

## KABUL VE ONAY

Recep Ali AY tarafından hazırlanan “Doğal Dil İşleme ve Kümeleme Yöntemleri ile Müşteri Yorumlarının İncelenmesi” başlıklı bu çalışma, 16 Mart 2021 tarihinde yapılan savunma sınavı sonucunda başarılı bulunarak jürimiz tarafından Yüksek Lisans/Doktora Tezi olarak kabul edilmiştir.

**Jüri Başkanı:** **Prof. Dr. Mehmet Reşit TOLUN** \_\_\_\_\_  
Konya Gıda ve Tarım Üniversitesi

**Tez Danışmanı:** **Dr. Öğr. Üyesi Semih YUMUŞAK** \_\_\_\_\_  
KTO Karatay Üniversitesi

**Jüri Üyesi:** **Dr. Öğr. Üyesi Ali Osman ÇIBIKDİKEN** \_\_\_\_\_  
KTO Karatay Üniversitesi

Jüri tarafından kabul edilen bu çalışmanın Yüksek Lisans/Doktora Tezi olması için gerekli şartları yerine getirdiğini onaylıyorum.

\_\_\_\_\_  
**Prof. Dr. Hüseyin Bekir YILDIZ**  
Enstitü Müdürü

## BİLDİRİM

Enstitü tarafından onaylanan Yüksek Lisans/Doktora tezimin tamamını veya herhangi bir kısmını basılı veya dijital biçimde arşivleme ve aşağıda belirtilen koşullar dahilinde erişime açma iznini KTO Karatay Üniversitesine verdiğimi bildiririm. Bu izinle, Üniversiteye verilen kullanım hakları dışındaki tüm fikri mülkiyet haklarım bende kalacak ve gelecekteki çalışmalar (makale, kitap, lisans, patent vb.) için tezimin tamamının veya bir bölümünün kullanım hakları yalnızca bana ait olacaktır.

Tezimin bütünüyle kendi çalışmam olduğunu, başkalarının haklarını ihlal etmediğimi ve tezimin tek yetkili sahibi olduğumu beyan ve taahhüt ederim. Telif hakkı bulunan ve sahiplerinden yazılı izinle kullanılması zorunlu olan kaynakları, yazılı izin alarak kullandığımı ve istenildiğinde izinlerin suretlerini Üniversiteye teslim etmeyi taahhüt ederim.

Yükseköğretim Kurulu tarafından yayımlanan “Lisansüstü Tezlerin Elektronik Ortamda Toplanması, Düzenlenmesi ve Erişime Açılmasına İlişkin Yönerge” kapsamında, tezim, aşağıda belirtilen koşullar haricince, YÖK Ulusal Tez Merkezi ve KTO Karatay Üniversitesi Açık Erişim Sisteminde erişime açılır.

Enstitü / Fakülte Yönetim Kurulu kararı ile tezimin erişime açılması mezuniyet tarihimden itibaren 2 yıl ertelenmiştir.<sup>1</sup>

Enstitü / Fakülte Yönetim Kurulunun gerekçeli kararı ile tezimin erişime açılması mezuniyet tarihimden itibaren ... ay ertelenmiştir.<sup>2</sup>

Tezimle ilgili gizlilik kararı verilmiştir.<sup>34</sup>

16 Mart 2021

---

**Recep Ali AY**

---

<sup>1</sup> MADDE 6(1) Lisansüstü teze ilgili patent başvurusu yapılması veya patent alma sürecinin devam etmesi durumunda, tez danışmanının önerisi ve enstitü anabilim dalının uygun görüşü üzerine enstitü veya fakülte yönetim kurulu iki yıl süre ile tezin erişime açılmasının ertelenmesine karar verebilir.

<sup>2</sup> MADDE 6(2) Yeni teknik, materyal ve metotların kullanıldığı, henüz makaleye dönüşmemiş veya patent gibi yöntemlerle korunmamış ve internetten paylaşılması durumunda 3. şahıslara veya kurumlara haksız kazanç imkanı oluşturabilecek bilgi ve bulguları içeren tezler hakkında tez danışmanının önerisi ve enstitü anabilim dalının uygun görüşü üzerine enstitü veya fakülte yönetim kurulunun gerekçeli kararı ile altı ayı aşmamak üzere tezin erişime açılması engellenebilir.

<sup>3</sup> MADDE 7(1) Ulusal çıkarları veya güvenliği ilgilendiren, emniyet, istihbarat, savunma ve güvenlik, sağlık vb. konulara ilişkin lisansüstü tezlerle ilgili gizlilik kararı, tezin yapıldığı kurum tarafından verilir. Kurum ve kuruluşlarla yapılan işbirliği protokolü çerçevesinde hazırlanan lisansüstü tezlere ilişkin gizlilik kararı ise, ilgili kurum ve kuruluşun önerisi ile enstitü veya fakültenin uygun görüşü üzerine üniversite yönetim kurulu tarafından verilir. Gizlilik kararı verilen tezler Yükseköğretim Kuruluna bildirilir.

<sup>4</sup> MADDE 7(2) Gizlilik kararı verilen tezler gizlilik süresince enstitü veya fakülte tarafından gizlilik kuralları çerçevesinde muhafaza edilir, gizlilik kararının kaldırılması halinde Tez Otomasyon Sistemine yüklenir.

## ETİK BEYAN

KTO Karatay Üniversitesi Lisansüstü Eğitim Enstitüsü Tez Hazırlama ve Yazım Kurallarına uygun olarak Dr. Öğr. Üyesi Semih YUMUŞAK danışmanlığında tarafımdan üretilen bu tez/proje çalışmasında; sunduğum tüm veri, enformasyon, bilgi ve belgeleri bilimsel etik kuralları çerçevesinde elde ettiğimi, tüm değerlendirme, analiz, bulgu ve sonuçları bilimsel usullere uygun olarak sunduğumu, tez/proje çalışmasında yararlandığım kaynakların tümüne bilimsel normlara uygun biçimde atıfta bulunarak kaynak gösterdiğimi, tezimin/projemin kaynak gösterilen durumlar dışında özgün olduğunu bildirir, aksi bir durumda aleyhime doğabilecek tüm hak kayıplarını kabullendiğimi beyan eder

16 Mart 2021

---

**Recep Ali AY**

## TEŐEKKÖR

Bu alıőmanın yűrűtűlmesi sırasında desteęini esirgemeyen danıőmanın Dr. Őęr. Ŭyesi Semih YUMUŐAK'a, aileme ve desteęini esirgemeyen herkese teőekkűr ederim.

16 Mart 2021

Recep Ali AY

## ÖZET

Recep Ali AY

Doğal Dil İşleme ve Kümeleme Yöntemleri ile Müşteri Yorumlarının İncelenmesi

Yüksek Lisans

Konya, 2021

Alışverişte internet kullanımının artması ile birlikte metin içerikli müşteri incelemeleri her geçen gün artmaktadır. Müşteri incelemeleri, ürün alımlarında ve hizmet alımlarında dikkate alınmaktadır. Müşteri karar verme süreçlerinde etkili olan müşteri incelemeleri şirketler için oldukça önemlidir. Ürün satışlarına direk etkisi düşünüldüğünde, şirketler sosyal medya ve internet yorumlarını cevaplamak ve yönetmek için istihdam sağlamaktadır. Ayrıca, müşteri yorumları yanıtıcı veya yönlendirici de olabilmekte, çeşitli yazılımsal araçlar ile manipüle edilebilmektedir. Makine öğrenmesi kullanarak benzer yorumların tespit edilmesi amaçlanan bu tez çalışmasında, Türkiye otelleri ile ilgili otel yorumları Python programlama dilinde geliştirilen bir araç ile otomatik olarak toplanarak bir veri kümesi oluşturulmuştur. Oluşturulan veri kümesi metin kümeleme yöntemi kullanılarak tez çalışmasında açıklanan yöntemlerle tasniflenmiştir. Bu tasnif etmenin sonucunda, birbirine benzer olan birçok yorumun farklı zamanlarda ve farklı işletmeler için oluşturulduğu tespit edilmiş, benzerlikler tespit edilen kümeler ve farklı yorum örnekleri tablolar halinde sunulmuştur.

### **Anahtar Kelimeler**

Otel Yorumları, Makine Öğrenmesi, Müşteri Memnuniyeti

## **ABSTRACT**

Recep Ali AY

Analysis of Customer Reviews by Using Natural Language Processing and Clustering

Methods

Master's D. Thesis

Konya, 2021

With the increase of the internet usage in shopping, textual customer reviews are increasing day by day. Customer reviews are taken into consideration in product purchases and service purchases. Customer reviews, which are effective in customer decision-making processes, are critically important for companies. Given its direct impact on product sales, companies employ people to respond and manage the social media and the internet reviews. In addition, customer reviews can be misleading or directive, and can be manipulated with various software tools. In this thesis, it is aimed to identify similar customer reviews using the machine learning methods, The Turkish hotels with hotel reviews were collected with a tool developed in Python, and a data set has been created. The data set created was clustered using the methods explained in the thesis study using the text clustering methods. As a result of this clustering, it was realized that many similar reviews were created at different times and for different businesses. In order to visualize and exemplify the results, clusters with similarities and different examples of reviews were presented in tables.

### **Keywords**

Hotel Reviews, Machine Learning, Customer Satisfaction



## İÇİNDEKİLER

KABUL VE ONAY .....	i
BİLDİRİM .....	ii
ETİK BEYAN.....	iii
TEŞEKKÜR.....	iv
ÖZET.....	ii
ABSTRACT.....	iii
İÇİNDEKİLER .....	iv
TABLolar DİZİNİ .....	v
ŞEKİLLER DİZİNİ.....	vii
KISALTMALAR DİZİNİ.....	viii
1. GİRİŞ .....	1
2. LİTERATÜR ÖZETİ.....	3
3. METODOLOJİ .....	8
3.1. DBSCAN Algoritması.....	8
3.2. Veri Toplama.....	12
3.3. Verilerin Analizi.....	17
4. SONUÇ .....	47
KAYNAKLAR .....	48
ÖZGEÇMİŞ .....	50

## TABLolar DİZİNİ

Tablo 1. Veritabanı Modeli .....	13
Tablo 2. Şehirlerdeki yorum sayısı .....	16
Tablo 3. En fazla yoruma sahip oteller .....	16
Tablo 4. Parça parça küme sayısı .....	21
Tablo 5. Tüm veri setinin küme sayıları .....	22
Tablo 6. Tüm verilerdeki kümelenemeyen yorum sayıları .....	22
Tablo 7. EPS 0.6'ya göre en fazla yorum içeren 20 küme .....	23
Tablo 8. EPS 0.7'ye göre en fazla yorum içeren 20 küme .....	24
Tablo 9. EPS 0.8'e göre en fazla yorum içeren 20 küme .....	25
Tablo 10. EPS 0.9'a göre en fazla yorum içeren 20 küme .....	27
Tablo 11. EPS 1.0'e göre en fazla yorum içeren 8 küme .....	28
Tablo 12. Kümeleme örnekleri .....	29
Tablo 13. Kümeleme örnekleri .....	30
Tablo 14. Kümeleme örnekleri .....	30
Tablo 15. Kümeleme örnekleri .....	31
Tablo 16. Kümeleme örnekleri .....	31
Tablo 17. Kümeleme örnekleri .....	32
Tablo 18. Kümeleme örnekleri .....	32
Tablo 19. Kümeleme örnekleri .....	33
Tablo 20. Kümeleme örnekleri .....	33
Tablo 21. Kümeleme örnekleri .....	35
Tablo 22. Kümeleme örnekleri .....	36
Tablo 23. Kümeleme örnekleri .....	38
Tablo 24. Kümeleme örnekleri .....	39
Tablo 25. Kümeleme örnekleri .....	40
Tablo 26. Kümeleme örnekleri .....	41
Tablo 27. Kümeleme örnekleri .....	42
Tablo 28. Kümeleme örnekleri .....	43
Tablo 29. Kümeleme örnekleri .....	43
Tablo 30. Kümeleme örnekleri .....	44
Tablo 31. Kümeleme örnekleri .....	45
Tablo 32. Kümeleme örnekleri .....	45

Tablo 33. Kümeleme örnekleri .....	46
Tablo 34. Kümeleme örnekleri .....	46

## ŞEKİLLER DİZİNİ

Şekil 1. Kümeleme örnekleri.....	8
Şekil 2. Komşuluk yarı çapı .....	9
Şekil 3. Veri kümesi dağılımı.....	10
Şekil 4. Noktaların kümelenmesi .....	11
Şekil 5. Veri toplama yazılımı: kütüphaneler .....	13
Şekil 6. Veri toplama yazılımı: şehir listesi .....	13
Şekil 7. Veri toplama yazılımı: veri tabanı bağlantısı.....	14
Şekil 8. Veri toplama yazılımı: otel tespiti.....	14
Şekil 9. Veri toplama yazılımı: oteller .....	14
Şekil 10. Veri toplama yazılımı: yorumları tespiti.....	15
Şekil 11. Analiz yazılımı: kütüphaneler.....	18
Şekil 12. Analiz yazılımı: DBSCAN EPS parametreleri .....	18
Şekil 13. Analiz yazılımı: mysql bağlantısı ve dataframe.....	19
Şekil 14. Analiz yazılımı: metin ön işleme .....	19
Şekil 15. Analiz yazılımı: TFIDF dönüşümü.....	20
Şekil 16. Analiz yazılımı: DBSCAN analizi.....	20
Şekil 17. EPS 0.6'ya göre en fazla yorum içeren 20 küme .....	24
Şekil 18. EPS 0.7'ye göre en fazla yorum içeren 20 küme .....	25
Şekil 19. EPS 0.8'e göre en fazla yorum içeren 20 küme .....	26
Şekil 20. EPS 0.9'e göre en fazla yorum içeren 20 küme .....	28
Şekil 21. EPS 1.0'e göre en fazla yorum içeren 8 küme .....	29

## KISALTMALAR DİZİNİ

<b>Kısaltma</b>	<b>Açıklama</b>
CNN	Convolutional Neural Networ
DBSCAN	Density-based spatial clustering of applications with noise
DDİ	Doğal Dil İşleme
EM	Ensemble Methods
K-NN	K-Nearest Neighbor
LR	Logistic Regression
ME	Maximum Entropy
NB	Naive Bayes
NLP	Natural Language Processing
RF	Random Forest
SQL	Structured Query Language
SVM	Support Vector Machine
TFIDF	Term Frequency–Inverse Document Frequency

## 1. GİRİŞ

Bryan Caplan Marketing (2020)'de yayımlanan verilere göre, alışveriş yapanların yaklaşık yüzde 95'i, bir satın alma işlemi gerçekleştirmeden önce çevrimiçi yorumları okumaktadır. Tüketicilerin yüzde 93'ü, yerel bir işletmenin iyi mi yoksa kötü mü olduğunu belirlemek için yorumları kullanmaktadır. Müşterilerin yüzde 72'si yorumları okumadan bir işlem yapmamaktadır. Potansiyel müşteriler, yerel bir işletmeye güvenileceklerini hissetmeden önce ortalama 10 çevrimiçi yorum okumaya ihtiyaç duymaktadır. Tüketicilerin yüzde 73'ü, olumlu yorumları okuduktan sonra yerel bir işletmeye daha çok güvenmektedir. Potansiyel müşteriler, bir işletmenin yıldız puanına güvenmeden önce ortalama 40 çevrimiçi yorum görmek istemektedir.

İşletmelerin rakipleri ile rekabet edebilmesi için müşterilerden gelen geri dönüşleri başarılı bir şekilde analiz ederek aksiyon almaları gerekmektedir. Geleneksel yöntemlerle veri toplayıp analiz etmek işletmelere yüksek maliyetlere getirmektedir. Ama internetin aktif kullanılmasıyla; çevrimiçi deneyimler, şikayetler ve öneriler işletmelerin ucuz ve çok fazla veri toplamalarına olanak vermektedir. İnternet üzerinden toplanan verilerin insan gücüyle analiz edilmesi uzun zaman alacaktır. Ayrıca her geçen gün toplanan verilere yeni verileri de eklediğimiz zaman insan gücüyle analiz yapılması imkansız hale gelecektir. İnternet üzerinde oluşan yüksek hacimli müşteri incelemeleri analiz edilerek anlamlı bilgiler elde edilmesi gerekmektedir.

Bilişim teknolojilerindeki gelişmeler ile birlikte bilgisayarlı dil bilimi çalışmaları son derece önem kazanmış, daha doğru ve hızlı sonuçlar elde edilmiştir. Doğal Dil İşleme yöntemleri insan ile bilgisayar etkileşiminde doğal diller kullanılması amacıyla başlamıştır. Bilgisayar tarafından veri kümesi üzerinde yapılan çalışma sonuçlarının insanlara doğal dil ile aktarılmasına “Konuşma”, insan tarafından konuşulanların bilgisayara girdi olarak işlenmesi ise “Konuşmayı Anlama” olarak adlandırılmıştır (Adalı, 2012).

Dil bir ulusu oluşturan temel yapı taşlarından bir tanesidir. Dünyada 7 bin üzeri dil çeşidi bulunmaktadır. Bu çeşitlilik doğal dil işleme alanında yapılan yaklaşımlar, metodolojiler, kaynaklar ve özellikle geliştirilen araçlar belirli dil ile sınırlı kalmakta ve diğer dillere transfer edilmesi mümkün değildir. Doğal Dil İşleme (DDİ) çalışmaları genellikle İngilizce metinler üzerinde yapılmaktadır. Bunun sebebi diğer diller için geliştirilmekte

olan araçlar İngilizce dilinde kullanılan araçlara göre doğruluk değerlerinin daha düşük olmasıdır.

İstatistiksel makine çevirisi yaklaşımında dil bilgisi açısından iki farklı hata türleri kontrol edilmektedir. Metin içerisinde kullanılan kelimenin; hatalı seçilmesi veya yanlış yerde kullanılmasıdır. Sinirsel ağların sürekli uzayda kelime ve ifadeleri modelleme ve doğrusal olmayan ilişkileri yakalama yeteneği, daha iyi genelleme ve daha doğru dilbilgisi hatası düzeltilmesi yapmalarını sağlamaktadır (Chollampatt vd., 2016) .

Morinaga vd. (2002), kendi ve rakiplerinin ürünlerinin itibarının bilinmesi pazarlama ve müşteri ilişkilerinin yönetilmesi için önemli olduğu bilgisini vermektedir. Liu vd. (2005), ürünlerin çevrimiçi müşteri incelemelerine odaklanmaktadır. Müşteri görüşlerini toplamak için internetin mükemmel bir kaynak haline geldiğini göstermektedir. Birçok ürün sayfası, forumlar, tartışma grupları ve bloglar müşteri incelemelerini içermektedir. Rakip ürünlerin tüketici görüşlerini analiz etmek ve karşılaştırmak için, Opinion Observer adlı bir prototip sistemi de uygulamaktır. Sistem görselleştirmesine bakıldığında kullanıcı her bir ürünün tüketicilerin kafasındaki çeşitli ürün özellikleri açısından güçlü ve zayıf yönlerini açıkça görebilmelerine olanak sağlamaktadır.

(Dave vd., 2003), metin madenciliği teknikleri kullanılarak belirlenen bir internet sayfası üzerinden ürün ile ilgili müşteri yorumları çekilerek ürün özelliklerine göre analizi yapılmıştır. Analizde yapılan yorum sayısının artması ürün özelliklerinin doğruluğu ve müşteri memnuniyeti ile ilgili yol gösterdiği tespit edilmiştir.

Bu tez çalışmasında, makine öğrenmesi ile Türkçe otel yorumları incelenerek manipülasyona neden olan yorumların tespit edilmesi amaçlanmaktadır. Çalışmada kümeleme algoritmalarından DBSCAN algoritması kullanılarak veri seti içerisinde bulunan manipülasyon olarak değerlendirilebilecek benzer yorumların tespit edilmesi amaçlanmıştır.

Geliştirilen bir yazılım sayesinde Türkçe otel yorumları barındıran sağlayıcılardan 81 ilde bulunan 2.623 otel ait 287.463 adet veri toplanmış ve anonimleştirilmiştir. Toplanan verilerde öznitelik olarak; şehir, otel adı, yorum, tarih, puan değerleri bulunmaktadır.

## 2. LİTERATÜR ÖZETİ

Hu vd. (2012) inceleme sahtekarlığını satıcıların, yayıncıların, yazarların veya herhangi bir üçüncü tarafın çevrimiçi incelemeleri izlemesi ve ürün satışlarını artırmak için gerçek müşteriler olarak gerçek olmayan çevrimiçi incelemeler yayınlaması olarak tanımlamaktadır.

Seyahat edenlerin incelemeleri sınırlı etki yaratmaktadır. Bu etkiyi arttırmak için otel yöneticileri misafirleri yalnızca yorum yazmaya teşvik etmek yerine müşterilerin kaldıkları zaman içerisinde nelerden memnun kaldıklarını okunabilir, doğru ve anlaşılır bir hikâye yazmaya yönlendirmeleri önerilmektedir (Raguseo vd., 2017).

Çevrimiçi incelemelerde belirtilen öznel özelliklerin farklı etkileri vardır. Tüketici talebiyle karşılaştırıldığında genel müşteri memnuniyeti, aynı zamanda tüketicilerin paylaşım düzeyine de bağlıdır. İşlem maliyetleri, özellikle bilgi arama ve edinme maliyetleri, paylaşım ekonomisinde tüketici satın alma kararlarını etkilemede önemli bir rol oynamaktadır. Somut özellikler için tüketicilerin, ev sahipleriyle doğrudan iletişim kurarak veya konaklama web sayfasındaki listeler ve açıklamalardan daha doğrudan bilgi edinme eğiliminde oldukları sonucuna varılmıştır Xu (2020).

Yetkin vd. (2016) Alanya’da yaptıkları bir çalışmada konaklama işletme yöneticilerinin otel değerlendirme sistemlerinden haberdar olduğu ve aynı zamanda düzenli bir şekilde inceleme yaptıkları belirtmektedirler. Olumsuz yorumların bedava bilgi kaynağı olduğunu ve hizmetlerin geliştirilmesinde etkili olduğu düşünülmektedir.

Mevcut çalışmalar çevrimiçi incelemelerde duyguların analizinde etkili olurken işletmelerin yanıt stratejilerine çok az ışık tutmaktadır. Chang vd. (2020) tarafından gerçekleştirilen Doğal Dil İşleme ile yapılan çalışmada otel temsilcilerinin incelemelere verilen yanıtları öncelikli sıraya sokmasına olanak vermektedir.

Chang vd. (2019) tarafından yapılan başka bir çalışma tek bir veri kaynağının otel puanlarının ve yorumlarının gerçek değeri keşfetmek için yeterli olmadığını göstermektedir. Birden fazla yerden alınan veri işlendiğinde her bir otelin hizmetini iyileştirmek ve pazarlama fırsatlarını keşfetmek için kullanılabilmesi tespit edilmiştir. Mobil cihazlar pazarlama için giderek daha fazla kullanılmaktadır. Sahte ve ücretli çevrimiçi incelemeler hakkında artan endişeler otellerin manipüle edildiğini göstermektedir. Ma vd. (2018), derin öğrenme modelleri, büyük veriyi işleme gücü ve



giderek daha karmaşık hale gelen algoritmaların kombinasyonunu bir araya getirerek birçok uygulama alanları için araştırma metodolojisini zenginleştirmektedir. Çalışmada, internette kullanıcı tarafından oluşturulan içeriklerin artmasıyla, otelcilik ve turizmde uzun süredir devam eden ve ortaya çıkan araştırma sorunlarının çoğu derin öğrenme modelleri ile birçok uygulama senaryolarının anlaşılmasına yardımcı olacağı belirtilmektedir.

Ku vd. (2019), gezginlerin faydalı yorumları elde edebilmeleri için; seyahat web sitelerinde yalnızca yorum tarihine ve puanına göre filtre uygulaması yapabildiğinden, yorumları incelemek için çok fazla zaman harcamaları gerekmektedir. Gezginler için seyahat bölgesi, türü ve zamana göre özelleştirilmiş faydalı incelemeler öneren akıllı bir öneri sistemi oluşturulmuştur. Gezginler için en uygun incelemelerin otomatik olarak filtrelenmesi harcanan zamanı azaltmakta ve böylece kullanıcıların web hizmetlerini kullanma şansını artırmaktadır.

Hlee vd. (2021) tarafından gerçekleştirilen çalışmada, tüketicilerin karar vermek için büyük ölçüde çevrimiçi incelemelere güvenmekte olduğu belirtilmektedir. Ayrıca çevrimiçi inceleme toplulukları ile tüketiciler artık çok sayıda gerçek inceleme arasından sahte incelemelerin nasıl etkili bir şekilde tespit edileceği konusunda ciddi bir sorunla karşı karşıya olduğu belirtilmiştir. Turizm ve konaklama endüstrilerinde güvenilir bir çevrimiçi pazarlama ortamı geliştirmek için bazı pratik çıkarımlar tespit edilmiştir. Bu çıkarımlara göre ilk olarak, bir işletmenin çevrimiçi sıralaması dikkate alınmalıdır. İkinci olarak, çevrimiçi pazarlamacıların sahte incelemeleri tespit etmek için aşırı puanlara sahip incelemelerin duyarlılık analizini benimseyen daha iyi bir filtreleme sistemi oluşturmaları önerilmektedir. Son olarak, birçok sahte incelemeyi araştırıp ortaya çıkarmak, turizm ve konaklama pazarlamacıları ile sahiplerinin sahte incelemeler oluşturmanın çevrimiçi ortamın ve çevrimiçi pazarın değerini yok ettiğinin farkına varmaları sağlanmalıdır.

Nilashi vd. (2018), sosyal medya araçlarının, seyahat planlama sürecinde ve müşterilerin seyahat tercihlerinde önemli bir rol oynadığını tespit etmiştir. Müşterilerin turizm ürünleri hakkındaki geri bildirimlerini ve yorumlarını paylaşmaları için farklı kanallara yönlendirilmiştir. Bu incelemeler, seyahatten önce otel rezervasyonu yapmak isteyen

yolcular için önemlidir. Gezginlerin otellerle ilgili yorumlarını kullanmak için yeni bir yöntem geliştirerek onlara tercihlerine göre uyarlanmış uygun oteller önermektedir.

Ahmetođlu ve Resul (2020) tarafından yapılan alıřmaya gre,dil, insanların kendilerini ifade etmesi iin gl bir mekanizmadır. Makine ile insan etkileřimini sađlamak iin makinelerin bu mekanizmayı bilmeleri gerekmektedir. Makineler DDİ ile metinlerin sayısal karřılıklarını bulurlar. Veri seti ierisindeki kelimelerle oluřturulan kelime torbası ile rastgele oluřturulan kelime torbasının eđitim sonucunda sınıflandırma bařarısı karřılařtırılmıřtır. Veri seti iinden seilen kelimeler ile yapılan eđitimin rastgele oluřturulan kelimelerden daha bařarılı olduđu tespit edilmiřtir. Bunun sebebinin eđitim sırasındaki kelime torbasındaki kelimelerin anlam iliřkisinin yksek olmasından kaynaklandığı belirtilmiřtir.

Ku vd. (2019) tarafından, otel incelemelerinin ve yanıt srelerinin otomatik olarak toplanması iin bir ara geliřtirilmiřtir. Bu verilerin eřitli ynlerini grmek iin bir makine đrenmesi modeli oluřturulmuřtur. Otel yneticilerinin farklı yanıt stratejileri kullanarak olumlu ve olumsuz yorumlara yanıt verdiklerinde pratik karar vermelerine olanak sađlayacađı belirtilmiřtir. Oluřturulan model veri analizi hakkında ek bilgiler de sunmaktadır.

Abro vd. (2020) tarafından yapılan bir alıřmada restoran incelemelerini ynlerine ve kutuplarına gre sınıflandırmak iin metin sınıflandırma teknikleri uygulanmıřtır. Word2vec'in, TFIDF zellik mhendisliđi teknikleriyle temsil edilen bigram ile karřılařtırıldıđında grevler iin daha iyi sonular gsterdiđi tespit edilmiřtir. Ayrıca SVM algoritmasının, NB, LR, RF ve Ensemble algoritmalarının sonularına gre daha iyi olduđu tespit edilmiřtir. alıřmadan elde edilen sonuların pratik neme sahip olduđu belirtilmekte olup, gelecekte yapılacak alıřmalara ıřık tutacađı ngrlmektedir.

Lin vd. (2015), in otel incelemesine iliřkin deneysel bir duygu sınıflandırması alıřması yapmıřtır. Kullanılan veri seti word2vec modeli ile kelime ve kelime bekleri zerinde eđitimler yapılmıřtır. SVM, Logistic Regression (LR), Convolutional, Neural Network (CNN), ensemble methods(EM), Naive Bayes (NB) ve Maximum Entropy (ME) đrenme yntemleri zerinde alıřmalar yapılmıřtır. Duygu analizi iin en az 60 bin yorumun bulunması gerekmektedir. Kullanılan veri arttıka performansın da arttıđı gzlenmiřtir. Kullanılan LR, SVM ve CNN đrenme yntemlerinin iyi performans gsterdiđi tespit

edilmiştir. Bu öğrenme yöntemleriyle ilgili olarak, SVM en iyi şekilde çalışmakta olduğu, CNN beklendiği kadar iyi çalışmasa da yine de NB ve ME'den daha iyi performansa sahip olduğu anlaşılmıştır.

Dey vd. (2016) tarafından yapılan çalışmada otel ve film incelemeleri duyarlılık sınıflandırması performansını doğruluk, kesinlik ve hatırlama açısından değerlendirilmiştir. Film incelemelerinde Naive Bayes'in % 80'in üzerinde doğruluk verdiği bildirilmiştir. K-NN yaklaşımının Naive Bayes'e göre performansının düşük olduğu ve daha iyi sonuç vermediği tespit edilmiştir. Otel incelemeleri için her iki sınıflandırıcının doğruluk seviyelerinin çok düşük olup, benzer sonuçlar verdiği belirtilmiştir. Bu nedenle Naive Bayes sınıflandırıcısının film incelemelerini analiz etmek için başarıyla kullanılabileceği önerilmiştir.

Fazzolari vd. (2017), bir inceleme metni ile ilişkilendirilen puanın hatalı olabileceğinden yola çıkarak bir çalışmagerçekleştirmiştir. Tripadvisor'dan 160 bin adet otel incelemesi alınarak içerikteki ifade edilen negatif ve pozitif kutupluluğa göre büyük bir veri kümesi oluşturulmuştur. Kullanılan Naive Bayes, Parça Algoritması ve SVM sınıflandırıcıları kullanılarak sınıflandırma yapılmıştır. Yapılan sınıflandırmada SVM %97 doğrulukla diğer sınıflandırıcılardan daha iyi sonuç elde edilmiştir.

Konaklama endüstrisi; dünyanın önde gelen işletmelerinden biri olmasına ve aynı zamanda ekonomisini istikrarlı bir şekilde artırmasına rağmen, büyük hacimli müşteri verilerinin doğru kullanımı konusunda çok az araştırma yapılmıştır. Konaklama sektörüne ait büyük verilerinin yoğunluğa dayalı kümeleme algoritmalarıyla analiz ederek veri kümeleme özelliklerine ilişkin çıkarımlarda bulunmaktadır. Elde edilen veri setine DBSCAN, OPTICS ve EnDBSCAN yoğunluğa dayalı algoritmaları kullanarak karşılaştırmalı performans analizini yapılmıştır. EnDBSCAN'ın gömülü kümeleri tanımlama açısından DBSCAN ve OPTICS'ten daha üstün performans gösterdiği tespit edilmiştir. Benzer şekilde, OPTICS, farklı veri kümeleri için bitişik iç içe geçmiş kümeleri tanımlamada DBSCAN'dan daha iyi performans göstermektedir. Ayrıca, konaklama endüstrisinin hem müşterilerin deneyimini hem de konaklama endüstrisinin gelirini iyileştirmesi için otomatikleştirilmiş öneri sistemlerinin etkinleştirilmesi için yeni kümeleme algoritmalarının geliştirilmesi gerektiği öngörülmektedir Bose vd. (2017) .

Barbado vd. (2019), çalışmasında sahte incelemelerin tespiti ele alınmıştır. Sosyal sitelerin analizine yönelik özellik çerçevesi önerilmiş olup, başka çalışmalar için kullanılmak üzere veri seti oluşturulmuştur. Çerçeve de inceleme merkezli ve kullanıcı merkezli iki ana özellik türünden oluşmaktadır. İnceleme merkezli özellikler yalnızca inceleme metni ile ilgilidir. Kullanıcı merkezli özellikler ise kullanıcının site içinde nasıl davrandığını göstermektedir. Kullanıcı merkezli özellikler kişisel, sosyal, gözden geçirme etkinliği ve güven olarak dört gruba ayrılır. Sahte incelemeleri yalnızca okuyarak tespit etmek hem insanlar hem de bilgisayarlar için zordur. Kullanıcıyla ilgili özelliklerin inceleme merkezlerine göre daha etkili olduğu tespit edilmiştir. Random Forest veya Ada Boost sınıflandırıcıları kullanıldığında maksimum %82 sonuca ulaşabilmektedir. TF-IDF gibi teknikler veya Word2vec gibi nöral öğrenme modelleri uygulanmasına rağmen, %60'ın altında sonuçlar elde edilmiştir. Sahte Özellik Çerçevesi kullanıcı merkezli incelemeleri dört alt gruba ayırma fikrinin başarılı olduğunu göstermektedir. Çünkü alt kümeler arasındaki özellikler bağımsızdır.

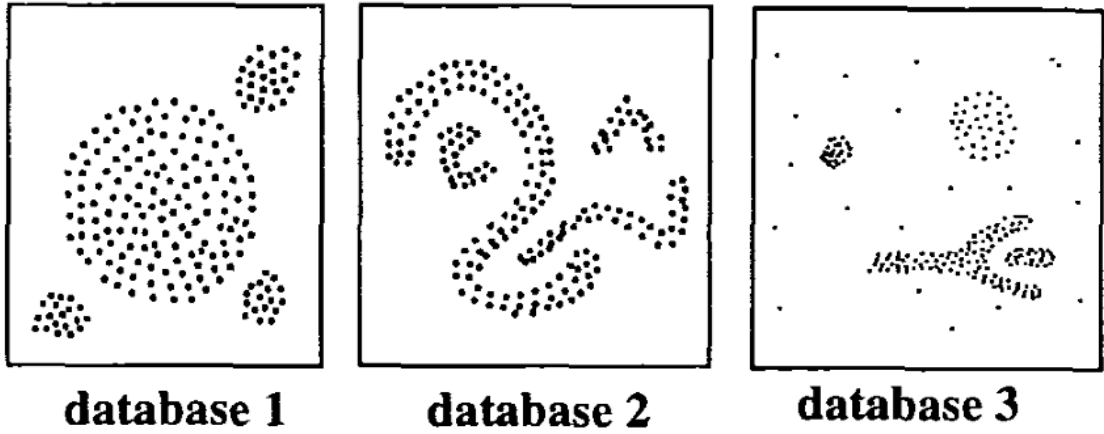
Carreón vd. (2019) tarafından Japon otellerin Çinli müşterilerinin ne istediğini anlamak amacıyla Çin portalı bir siteden Japon otellerinin incelemeleri elde edilmiştir. Makine öğrenmesi deneylerinde en yüksek performansı elde edebilmek için SVM'yi eğitirken hem pozitif hem de negatif anahtar kelimeleri kullanılmıştır. Japon otellerinin Çinli müşterilerinin fiyatla daha az ilgilendikleri otelin yakınındaki restoran, bahçeler veya parkları ulaşım kolaylıklarını dikkate aldıkları tespit edilmiştir.

### 3. METODOLOJİ

Çalışmanın bu bölümünde kullanılan kümeleme algoritması (DBSCAN), verilerin toplanması, verilerin analizi yöntemleri açıklanmaktadır.

#### 3.1. DBSCAN Algoritması

DBSCAN algoritması, Ester vd. (1996) tarafından ilk olarak önerilmiştir. Etkili bir yoğunluk temelli kümeleme yöntemidir. Diğer kümeleme yöntemleri ile karşılaştırıldığında, DBSCAN birkaç farklılaştırıcı özelliğe sahiptir. İlk olarak, verileri rastgele şekillere sahip kümelere bölebilir. Örneğin, tamamen başka bir küme ile çevrili kümeler bulabilir. İkinci olarak, DBSCAN'de önceden küme sayısını tanımlanmasına gerek yoktur. Üçüncüsü, veri kümesindeki noktaların sırasına duyarlıdır.

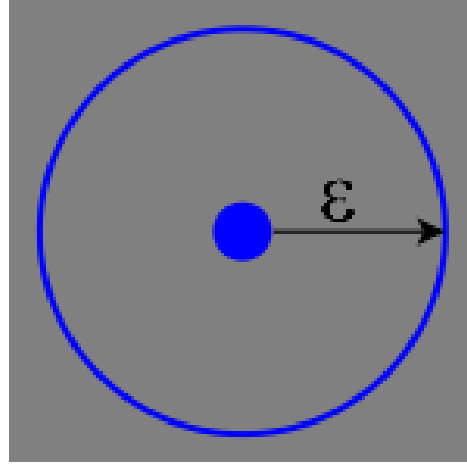


Şekil 1. Kümeleme örnekleri

(Ester vd., 1996)

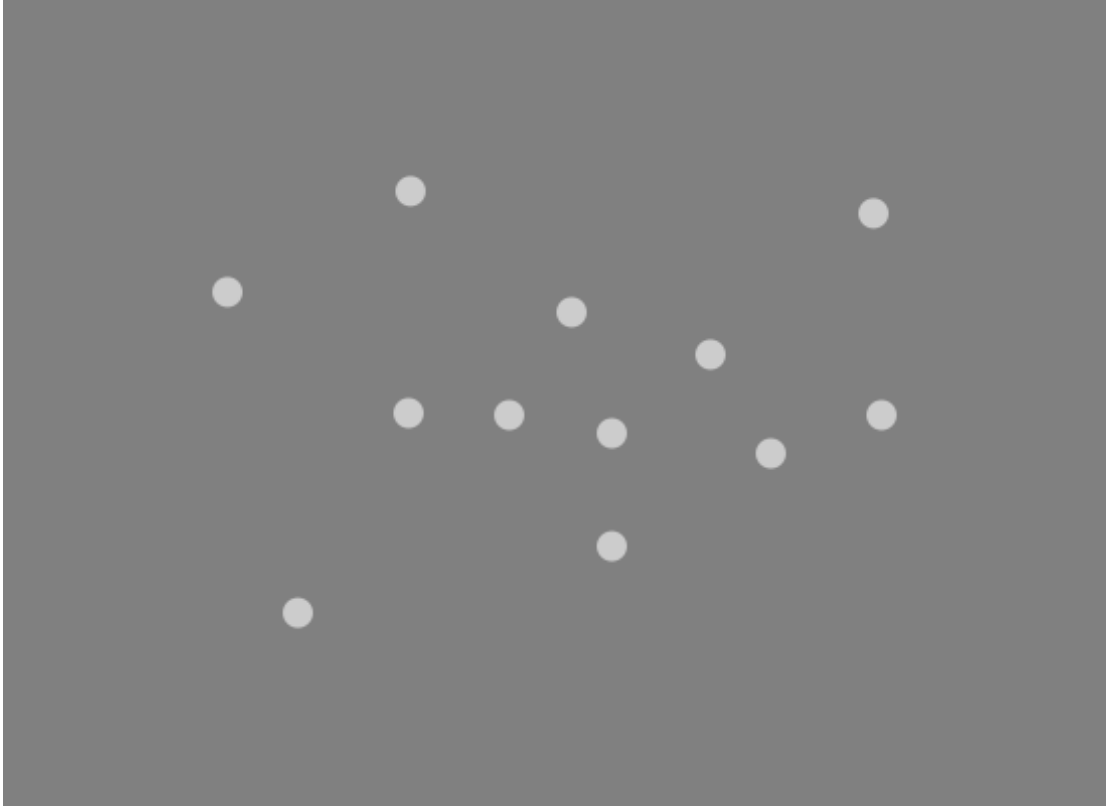
Şekil 1'de tasvir edilen örnek nokta kümelerine bakarken, bu kümelerin hiçbirine ait olmayan nokta kümelerini ve gürültü noktalarını kolayca ve açık bir şekilde tespit edilmektedir. Kümelerin tanınmasının ana nedeni, her kümenin içinde kümenin dışından önemli ölçüde daha yüksek olan tipik bir nokta yoğunluğuna sahip olmasıdır. Ayrıca, gürültü alanları içindeki yoğunluk, kümelerin herhangi birindeki yoğunluktan daha düşüktür. Temel fikir, bir kümenin her noktası için belirli bir yarıçapın komşuluğunun en azından minimum sayıda nokta içermesi gerektiğidir. Yani komşuluktaki yoğunluğun bir eşiği aşması gerekir. Bir komşuluğun şekli,  $dist(p, q)$  ile gösterilen iki nokta  $p$  ve  $q$  için bir uzaklık fonksiyonunun seçimi ile belirlenir.

DBSCAN algoritması girdi parametresi olarak iki deęişken kullanır: epsilon ve minPoints. Epsilon, yoğunluęu kontrol etmek için her veri noktası çevresinde oluşturulacak dairenin yarıçapıdır ve o veri noktasının çekirdek noktası olarak sınıflandırılması için bu daire içinde bulunması gereken minimum veri noktası sayısıdır. Şekil 2 de bir noktanın EPS deęeri gösterilmektedir.



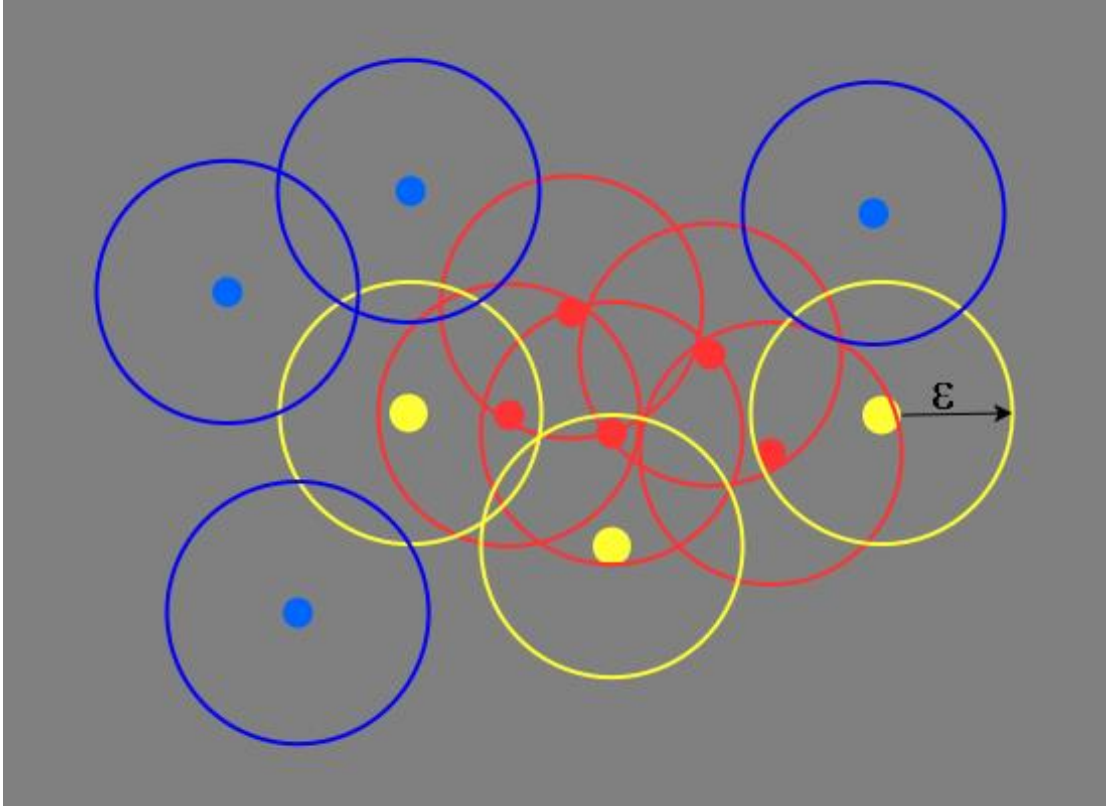
**Şekil 2. Komşuluk yarı çapı**

Şekil 3'te gri renkle temsil edilen bazı veri noktaları gösterilmektedir. DBSCAN, her veri noktasının etrafında epsilon yarıçaplı bir daire oluşturur ve bunları çekirdek noktası, sınır noktası ve gürültü olarak sınıflandırır. Bir veri noktası, etrafındaki daire en az "minPoints" sayıda nokta içeriyorsa bir çekirdek noktadır. Nokta sayısı minPoints'ten az ise, o zaman sınır noktası olarak sınıflandırılır ve epsilon yarıçapı içinde herhangi bir veri noktasının etrafında başka veri noktası yoksa gürültü olarak deęerlendirilir.



**Şekil 3. Veri kümesi dağılımı**

Şekil 4’te minPoints değeri 3 olacak şekilde DBCAN tarafından oluşturulan bir küme gösterilmektedir. Burada, her veri noktasının etrafında epsilon eşit yarıçaplı bir daire çizilmektedir. EPS ile minPoints parametreleri, uzamsal kümelerin oluşturulmasına yardımcı olmaktadır. Çemberde kendisi dahil en az 3 nokta bulunan tüm veri noktaları kırmızı renkle temsil edilmekte ve çekirdek noktalar olarak kabul edilmektedir. Çemberde kendisi dahil 3’ten küçük ancak 1’den büyük olan tüm veri noktaları sınır noktaları olarak adlandırılmaktadır. Sarı renk ile sınır noktalar gösterilmektedir. Son olarak, daire içinde kendisinden başka hiçbir noktası olmayan veri noktaları gürültü olarak adlandırılıp, mavi renk ile gösterilmektedir.



**Şekil 4. Noktaların kümelenmesi**

Uzayda veri noktalarını bulmak için, DBSCAN öklid mesafesini kullanır, ancak başka yöntemler de kullanılabilir (coğrafi veriler için büyük daire mesafesi gibi). Aynı zamanda tüm veri kümesini bir kez taraması gerekirken, diğer algoritmalarda tarama işlemi birden çok kez yapılması gerekebilir.

**Avantajları:**

- Verilerdeki küme sayısının önceden belirtilmesine gerek yoktur.
- Rastgele şekilli kümeler bulabilir.
- Bir gürültü kavramına sahiptir ve aykırı değerlerin tespit edilmesine olanak verir.
- Yalnızca iki parametre kullanılır.
- Giriş veri kümesinin yalnızca bir kez taramasına ihtiyaç duyar.

**Dezavantajları:**

- Bir set üzerinde farklı yoğunluklara sahip kümeler olabilir. Bu durumda Epsilon değeri bir küme için uygun olurken, diğer küme için uygun olmayabilir.



- Büyük ölçekli veri setleri üzerinde uzun sürebilir.
- Giriş parametrelerini kestirmek zordur.
- Bir veri setine yeni bir eleman eklendiğinde, bu elemanın hangi kümeye girdiğini bulmak kolay olmayabilir.
- Bir küme içinde başka bir küme yer alıyorsa, onu keşfetmek mümkün olmayabilir.

### 3.2. Veri Toplama

Veriler Türkçe yorum barındıran servis sağlayıcılarından elde edilmiştir. 81 ilde bulunan 2.623 otel ait 287.463 adet veri toplanmıştır. Öznitelik olarak;

- Şehir,
- Otel adı,
- Yorum,
- Tarih,
- Puan seçilmiştir.

Bilgiler, tez çalışması sürecinde Python programlama dili ile yazılan bir komut dosyası ile toplanmıştır. Komut dosyasında urllib ve BeautifulSoup kütüphanelerinden yararlanılmıştır. Urllib kütüphanesi sitelerden html kodlarını çekmekte kullanılmıştır. BeautifulSoup kütüphanesi çekilmiş html kodlarının içerisindeki verilerin elde edilmesinde kullanılmıştır. Programda öncelikle illere göre otellerin bulunması için bir döngü oluşturulmuştur. Burada bulunan otel isimlerine göre otellerin kaç sayfa yorum barındırdığı bilgisi elde edilmiştir. Elde edilen sayfa bilgisine göre BeautifulSoup kütüphanesi kullanılarak yorum, tarih, puan bilgisi alınmıştır. Veriler Mysql üzerinden oluşturulan bir tabloya şehir, otel adı, yorum, tarih ve puan bilgileri kaydedilmiştir. Veritabanında oluşturulan tabloya ait alanları ve veri türleri Tablo 1’de gösterilmiştir. Elde edilen verilerden en fazla yoruma sahip ilk 10 otel bilgileri Tablo 3’te gösterilmiştir.

**Tablo 1. Veritabanı Modeli**

Sıra	Alan Adı	Veri Türü	Açıklama
1	id	Int <i>Primary Key</i>	İndis Değeri
2	sehir	varchar(50)	Şehir
3	otel	varchar(100)	Otel Adı
4	puan	Text	Otel Puanı
5	tarikh	Date	Yorum Tarihi

Yorumların online olarak toplanması için yazılan komut dosyası ile otel yorumları toplanmıştır. Aşağıda komut dosyası parça parça açıklanmaktadır. Şekil 5'te kullanılan kütüphaneler gösterilmektedir.

```
1 from bs4 import BeautifulSoup
2 import urllib
3 from urllib.request import urlopen
4 import time
5 import re
6 import math
7 import numpy
8 import mysql.connector
9
```

**Şekil 5. Veri toplama yazılımı: kütüphaneler**

Şekil 6'da veri toplanacak site adresi alınmakta olup, site içerisinde aranacak şehir isimleri bir dizi olarak tanımlanmıştır

```
10 adres = "xxxxxxxxxxxxx.com"
11
12 sehirler=["Adana", "Adiyaman", "Afyonkarahisar", "Agri", "Aksaray", "Amasya",
13 "Ankara", "Antalya", "Ardahan", "Artvin", "Aydin",
14 "Balikesir", "Bartın", "Batman", "Bayburt", "Bilecik", "Bingol", "Bitlis", "
15 Bolu", "Burdur", "Bursa", "Canakkale", "Cankiri", "Corum",
16 "Denizli", "Diyarbakir", "Duzce", "Edirne", "Elazig", "Erzincan", "Erzurum", "
17 Eskisehir", "Gaziantep", "Giresun", "Gumushane", "Hakkari",
18 "Hatay", "Igdir", "Isparta", "Istanbul", "Izmir", "Kahramanmaras", "Karabuk",
19 "Karaman", "Kars", "Kastamonu", "Kayseri", "Kilis",
20 "Kirikkale", "Kirkklareli", "Kirsehir", "Kocaeli", "Konya", "Kutahya", "Malatya",
21 "Manisa", "Mardin", "Mersin", "Mugla", "Mus", "Nevsehir",
22 "Nigde", "Ordu", "Osmaniye", "Rize", "Sakarya", "Samsun", "Sanliurfa", "Siirt",
23 "Sinop", "Sivas", "Sirnak", "Tekirdag", "Tokat",
24 "Trabzon", "Tunceli", "Usak", "Van", "Yalova", "Yozgat", "Zonguldak"]
25
```

**Şekil 6. Veri toplama yazılımı: şehir listesi**

Şekil 7’de mysql bağlantısı sağlamak için kullanılan kod parçası yer almaktadır.

```
20
21 reObj=re.compile('\d+')
22
23 mydb = mysql.connector.connect(
24     host="xxxxxx",
25     user="xxxxx",
26     passwd="xxxxx",
27     database="xxxxx"
28 )
```

**Şekil 7. Veri toplama yazılımı: veri tabanı bağlantısı**

Şekil 8’de tarama yapılacak adres altında bulunan şehirlere ait otellerin kaç sayfada olduğunun tespit edildiği kod bloğu bulunmaktadır.

```
29 for sehir in sehirler:
30     # yol = sehir + '.txt'
31     # dosya = open(yol, 'w')
32     mycursor = mydb.cursor()
33     sql = "INSERT INTO data (sehir, otel, tarih, puan, yorum) VALUES (%s, %s, %s, %s, %s)"
34     sayfa = urllib.request.urlopen(adres+sehir)
35     soup = BeautifulSoup(sayfa, "html.parser")
36     oteller_raw = soup.findAll('div', attrs={"class":"items"})
37     sayfaNumarasi = soup.findAll('div', attrs={"class":"result"})
38     sayfaNumarasi = math.ceil(int(reObj.findall(str(sayfaNumarasi))[0])/12)
39
40     oteller = []
41
```

**Şekil 8. Veri toplama yazılımı: otel tespiti**

Şekil 9’da şehir altında tespit edilen otel adresleri, oteller dizisinde toplanmaktadır.

```
41
42 for otel in oteller_raw:
43     a = otel.find('a', href=True)
44     if (a != None):
45         oteller.append(a.get("href"))
46     for i in range(2, (sayfaNumarasi+1)):
47         sayfa = urllib.request.urlopen(adres+sehir+"/"+str(i))
48         soup = BeautifulSoup(sayfa, "html.parser")
49         oteller_raw = soup.findAll('div', attrs={"class":"items"})
50         for otel in oteller_raw:
51             a = otel.find('a', href=True)
52             if (a != None):
53                 oteller.append(a.get("href"))
54
```

**Şekil 9. Veri toplama yazılımı: oteller**

Şekil 10’da tespit edilen oteller içerisindeki yorumlar elde edilmeye çalışıldığı kod bloğu görülmektedir. Otele ait sayfalardaki HTML kodları BeautifulSoup kütüphanesi sayesinde veriler parçalanarak yorum, tarih ve puan değerleri elde edilmiştir. Elde edilen

verilerin daha sonra kümeleme algoritmasında kullanılmak üzere veritabanına kaydedilmesi sağlanmıştır.

```
54
55     try:
56         for otel in oteller:
57             url = adres+otel
58             otelSayfa = urllib.request.urlopen(url)
59             soup = BeautifulSoup(otelSayfa, "html.parser")
60             yorum_raw = soup.findAll('div', attrs={"class":"user-comment"})
61             yorumSayfaNumarasi = soup.findAll('div', attrs={"class":"result"}
62             )
63             yorumSayfisi = int(reObj.findall(str(yorumSayfaNumarasi))[0])
64             yorumSayfaSayfisi = math.ceil(yorumSayfisi/10)
65             for x in yorum_raw:
66                 yorum = x.find('div', attrs={'class':'header'})
67                 yorum = yorum.find('span').text
68                 tarih = str(x.find('div', attrs={'class':'date'}))
69                 puan = x.find('div', attrs={'class':'rating-text'})
70                 puan = puan.find('strong').text
71                 dosya.write(sehir+"<->" + otel[4:] + "<->" + tarih.replace(" ", "
72                 ").replace("\n", " ") + "<->" + str(puan) + "<->" + yorum.strip
73                 () + "\n")
74                 val = (sehir, otel[4:], tarih.replace(" ", "").replace("\n",
75                 " "), str(puan), yorum.strip())
76                 mycursor.execute(sql, val)
77                 mydb.commit()
78                 print "data"
79
80             for ii in range(2, (yorumSayfaSayfisi+1)):
81                 print(sehir+"--"+otel+"--"+str(ii))
82                 url = adres+otel+"/yorumlar-"+str(ii)
83                 otelSayfa = urllib.request.urlopen(url)
84                 soup = BeautifulSoup(otelSayfa, "html.parser")
85                 yorum_raw = soup.findAll('div', attrs={"class":"user-comment"}
86                 )
87                 for x in yorum_raw:
88                     yorum = x.find('div', attrs={'class':'header'})
89                     yorum = yorum.find('span').text
90                     tarih = str(x.find('div', attrs={'class':'date'}))
91                     puan = x.find('div', attrs={'class':'rating-text'})
92                     puan = puan.find('strong').text
93                     val = (sehir, otel[4:], tarih.replace(" ", "").replace("
94                     \n", " "), str(puan), yorum.strip())
95                     mycursor.execute(sql, val)
96                     mydb.commit()
97
98     except Exception as e:
99         pass
```

Şekil 10. Veri toplama yazılımı: yorumları tespiti

Toplamda, Türkçe yorum barındıran servis sağlayıcılarından 81 ilde bulunan 2.623 otel ait 287.463 adet veri toplanmıştır. En fazla yoruma sahip 10 şehir Tablo 2’de gösterilmektedir. Otel bazlı olarak incelendiğinde 0B525 hash koduna sahip otelin 11.814 adet yoruma sahip olduğu gözlenmektedir. En fazla yoruma sahip 10 otele ait bilgiler Tablo 3’de gösterilmektedir.

**Tablo 2. Şehirlerdeki yorum sayısı**

Sıra	Şehir	Yorum Sayısı
1	İstanbul	45731
2	İzmir	38631
3	Muğla	27225
4	Bursa	23020
5	Aydın	21824
6	Afyonkarahisar	16591
7	Ankara	16318
8	Nevşehir	12507
9	Çanakkale	10564
10	Eskişehir	9336
11	Antalya	6967
12	Bolu	6392

Tablo 2'deki şehirlere ait yorum sayıları veritabanında bulunan verileri içerisinde aşağıdaki sql komutu ile elde edilmiştir

```
SELECT sehir, COUNT(sehir) FROM data GROUP BY sehir ORDER BY COUNT(sehir) DESC LIMIT 12
```

**Tablo 3. En fazla yoruma sahip oteller**

Sıra	Otel Hash	Şehir	Yorum Sayısı
1	0B525	Muğla	11814
2	16DD0	İzmir	8435
3	59A6C	Aydın	5828
4	361A9	Afyonkarahisar	5666

5	00C5D	Antalya	4989
6	4231A	İzmir	3759
7	B8616	Aydın	3519
8	A224D	Sakarya	3445
9	C15BF	Muğla	3292
10	8C438	Bolu	3057

Tablo 3 içerisinde listelenen otellere ait yorum sayıları veritabanında bulunan verileri içerisinde aşağıdaki sql komutu ile elde edilmiştir

```
SELECT otel_hash, sehir, COUNT(otel_hash)FROM data GROUP BY otel ORDER BY COUNT(otel) DESC LIMIT 10
```

### **3.3. Verilerin Analizi**

Otel yorumları veri kümesi, Python programı dili üzerinde DBSCAN kümeleme algoritması kullanılarak kümeleme işlemine tabi tutulmuştur. Kümeleme işlemi sürecinde; numpy, pymysql, pandas, sklearn, re, nltk kütüphaneleri kullanılmıştır.

NumPy, bilimsel hesaplamalarda kullanılan bir matematik kütüphanesidir. Python dizilerine göre daha hızlı ve işlevseldir.

Pymysql kütüphanesi, mysql veritabanına bağlantı yapıp sorgu gönderilmesine olanak sağlamaktadır.

Pandas, numpy kütüphanesinin eksik kaldığı yerleri düzenlemektedir. Özellikle dizilerin homeojen yapıda olmasına gerek yoktur.

Sklearn, python programlama dili için yazılmış makine öğrenmesi kütüphanesidir.

Re, düzenli ifadeler için kullanılan kütüphanedir.

NLTK insan dili verileriyle çalışmak üzere Python programları oluşturmak için kullanılan bir kütüphanedir. Sınıflandırma, belirteç oluşturma, kök belirleme, etiketleme, ayrıştırma ve anlamsal akıl yürütme için metin işleme kitaplıklarıdır.

Şekil 11’de kullanılan kütüphane ve sınıflar bulunmaktadır.

```
1 import os
2 import numpy as np
3 import pymysql
4 from pandas import DataFrame
5 from sklearn.feature_extraction.text import TfidfVectorizer
6 from sklearn.preprocessing import StandardScaler
7 from sklearn.decomposition import PCA
8 import re
9 import nltk
10 nltk.download('stopwords')
11 from nltk.corpus import stopwords
12 from sklearn.cluster import DBSCAN
13 import matplotlib.pyplot as plt
14 from time import time, sleep
15 import pandas as pd
16 from nltk.stem import WordNetLemmatizer
17 import datetime
18
```

Şekil 11. Analiz yazılımı: kütüphaneler

Şekil 12’de DBSCAN algoritmasında kullanılan farklı girdi parametre değerleri için oluşturulan Epsilon değeri 0.6 ile 1.6 arasında 0.1 artış miktarına sahip sözlük dizisi bulunmaktadır.

```
19
20 epsValues = [{'columnName': 'epsTT6', 'value': 0.6},
21              {'columnName': 'epsTT7', 'value': 0.7},
22              {'columnName': 'epsTT8', 'value': 0.8},
23              {'columnName': 'epsTT9', 'value': 0.9},
24              {'columnName': 'epsTT10', 'value': 1.0},
25              {'columnName': 'epsTT11', 'value': 1.1},
26              {'columnName': 'epsTT12', 'value': 1.2},
27              {'columnName': 'epsTT13', 'value': 1.3},
28              {'columnName': 'epsTT14', 'value': 1.4},
29              {'columnName': 'epsTT15', 'value': 1.5},
30              {'columnName': 'epsTT16', 'value': 1.6}]
31
```

Şekil 12. Analiz yazılımı: DBSCAN EPS parametreleri

Şekil 13’te mysql veritabanı bağlantısı sağlayan kod bulunmaktadır. Bağlantı ile otel yorumları id, yorum ve puan değerleri alınarak, pandas kütüphanesinin *DataFrame* türüne dönüşümü sağlanmaktadır. Veri çekme işlemi tamamlandıktan sonra mysql bağlantısı kapatılmaktadır.

```

32
33 db = pymysql.connect("XXXXX", "XXXX", "XXXXXX", "XXXXX" )
34 cursor = db.cursor()
35 sql = "SELECT id, yorum, puan FROM data"
36 cursor.execute(sql)
37 data = DataFrame(cursor.fetchall(), columns = ['id', 'yorum', 'puan'])
38 db.close()
39 X, y = data.yorum, data.puan
40 documents = []
41

```

**Şekil 13. Analiz yazılımı: mysql bağlantısı ve dataframe**

Şekil 14'te DataFrame haline getirilen verilere tek tek metin ön işleme adımları uygulanma işlemleri gösterilmektedir. Bu adımlar;

- Tüm özel karakterlerin kaldırılması,
- Tüm tek karakterlerin kaldırılması,
- Baştaki tek karakterlerin kaldırılması,
- Birden fazla boşlukların bire indirilmesi,
- Metnin küçük harfe çevrilmesi,
- Lemmatizasyon işlemidir.

```

42
43 stemmer = WordNetLemmatizer()
44 for sen in range(0, len(X)):
45     document = re.sub(r'\W', ' ', str(X[sen]))
46     document = re.sub(r'\s+[a-zA-Z]\s+', ' ', document)
47     document = re.sub(r'\^[a-zA-Z]\s+', ' ', document)
48     document = re.sub(r'\s+', ' ', document, flags=re.I)
49     document = re.sub(r'^\b\s+', ' ', document)
50     document = document.lower()
51     document = document.split()
52     document = [stemmer.lemmatize(word) for word in document]
53     document = ' '.join(document)
54     documents.append(document)

```

**Şekil 14. Analiz yazılımı: metin ön işleme**

Şekil 15'de veri kümesinin ön işleme tamamlandıktan sonra veriyi makinenin anlayacağı hale getirmek için NLTK'nın stopwords fonksiyonu kullanılarak gereksiz kelimelerden temizlenmesi işlemi gösterilmektedir. Bu aşamada, gereksiz, sadece öznitelik çıkarımı yaparken önemli olmayan kelimelerin (ben, sen, de, da, ki, ile vs) çıkarılması işlemi uygulanmıştır. Gereksiz kelimelerden temizleme işleminden sonra verilerin makinenin anlayacağı sayısal değerlere dönüşümü yapılması gerekmektedir. Bu dönüşümde TFIDF vektörü kullanılmıştır.



```

56
57 tfidfconverter = TfidfVectorizer(stop_words=stopwords.words('turkish'), use_idf=True,
    max_features=10000, min_df=5)
58 X = tfidfconverter.fit_transform(documents).toarray()
59 pca = PCA(n_components=2)
60 X_pca = pca.fit_transform(X)
61 df = DataFrame(data, columns=['id', 'yorum', 'puan'])

```

**Şekil 15. Analiz yazılımı: TFIDF dönüşümü**

Şekil 16’da EPS adındaki sözlük dizisi değişkeninde tanımlanan değerler için ayrı ayrı DBSCAN ile kümeleme işlemi yapılmaktadır. Kümeleme işlemleri tüm veri kümesi için EPS değeri 0.6 ile 1.6 arasında 0.1 artış oranlarına göre ayrı ayrı oluşturularak küme numaraları veritabanında farklı sütunlar halinde kaydedilmiştir.

```

64 for epsValue in epsValues:
65     startDate = datetime.datetime.now()
66     startTime = time()
67     db = DBSCAN(eps=epsValue["value"], metric='euclidean', min_samples=2, n_jobs=-1,
        leaf_size=30).fit(X)
68     endTime = time()
69     endDate = datetime.datetime.now()
70     core_samples_mask = np.zeros_like(db.labels_, dtype=bool)
71     core_samples_mask[db.core_sample_indices_] = True
72     labels = db.labels_
73     n_clusters = len(set(labels)) - (1 if -1 in labels else 0)
74     unique_labels = set(labels)
75     colors = [plt.cm.Spectral(each)
        for each in np.linspace(0, 1, len(unique_labels))]
76
77     for k, col in zip(unique_labels, colors):
78         if k == -1:
79             col = [0, 0, 0, 1]
80
81         class_member_mask = (labels == k)
82
83         xy = X_pca[class_member_mask & core_samples_mask]
84         plt.plot(xy[:, 0], xy[:, 1], 'o', markerfacecolor=col,
85                 markeredgecolor='k',
86                 markersize=14)
87
88         xy = X_pca[class_member_mask & ~core_samples_mask]
89         plt.plot(xy[:, 0], xy[:, 1], 'o', markerfacecolor=col,
90                 markeredgecolor='k',
91                 markersize=6)
92
93     plt.title('number of clusters: '+str(n_clusters)+' , eps: '+str(epsValue['value'])+',
94             number of comments: '+str(len(db.labels_))+'')
95     plt.savefig("report/"+str(ofset)+"/"+str(epsValue['columnName'])+'.png')
96
97     db2 = pymysql.connect("XXXXX", "XXXXX", "XXXXXX", "XXXXXX")
98     cursor = db2.cursor()
99     for i in range(len(db.labels_)):
100         sql = "UPDATE data SET "+epsValue['columnName']+ "=" +str(db.labels_[i])+" WHERE id="+
            str(data.id[i])
101         cursor.execute(sql)
102         db2.commit()
103     db2.close()
104

```

**Şekil 16. Analiz yazılımı: DBSCAN analizi**

Kümeleme çalışması üç farklı şekilde gerçekleştirilmiştir. Yukarıda kullanılan yöntemler öncelikle her otel için ayrı kümeleme yapılacak şekilde uygulanmıştır. Verilerde yapılan incelemelerde otel bazlı kümelemenin güvenilir bir sonuç vermediği gözlemlendiğinden kümeleme işlemi genişletilmiştir. Aynı kümeleme işlemi veri kümesinin beş eşit parçaya

bölünmesiyle tekrar gerçekleştirilmiştir. Elde edilen sonuçlar incelendiğinde anlamlı bir sonuç olmadığı gözlemlenerek verilen yeniden analizi yapılmıştır. Yapılan son kümeleme çalışmasında tüm veriye tek parça olarak kümeleme işlemi uygulanmıştır.

Tüm veri kümesi üzerinde yapılan kümeleme çalışmasında komşuluk sorgularında kullanılacak EPS değeri 0.6 ile 1.0 arasındaki değerlerde kümeleme yapıldığı gözlenmiştir. 1.0'den büyük EPS değerlerine sahip kümeleme çalışmaları incelendiğinde yorumların hepsinin tek bir kümede toplandığı tespit edilmiştir. Tablo 4 içerisinde bütün otellere ait yorumların 5 eşit parçada yapılan kümeleme çalışmasının küme adetleri gösterilmektedir. Yorumların küme numaralarına göre inceleme yapıldığında manipülasyona yönelik karşılaştırma yapmanın yetersiz olduğu gözlenmiştir.

**Tablo 4. Parça parça küme sayısı**

Veri Kümesi	EPS 0,6	EPS 0,7	EPS 0,8	EPS 0,9	EPS 1,0	EPS 1,1	EPS 1,2	EPS 1,3	EPS 1,4	EPS 1,5	EPS 1,6
Grup 1	1105	959	849	689	25	1	1	1	1	1	1
Grup 2	1084	982	887	724	34	1	1	1	1	1	1
Grup 3	1140	1020	826	628	21	1	1	1	1	1	1
Grup 4	1127	987	836	640	17	1	1	1	1	1	1
Grup 5	8758	7927	6969	5701	1176	1	1	1	1	1	1

Tüm veri setinin kullanılarak yapılan kümeleme çalışması 81 ilde bulunan 2.623 otel ait 287.463 yorumun Tablo 5'te EPS değerlerine göre oluşturulan küme sayıları ve kümeleme süreleri gösterilmektedir. Tablo 6'da ise kümeleme dışı kalan yorum sayıları gösterilmektedir.

**Tablo 5. Tüm veri setinin küme sayıları**

Sıra	EPS Deęeri	Küme Sayısı	Kümeleme Süresi
1	0.6	57034	12:19:12
2	0.7	50998	15:07:51
3	0.8	43646	22:49:14
4	0.9	34518	33:46:01
5	1	6796	44:35:27
6	1.1	1	46:54:09
7	1.2	1	45:48:41
8	1.3	1	44:54:54
9	1.4	1	57:30:27
10	1.5	1	178:00:32
11	1.6	1	200:13:39

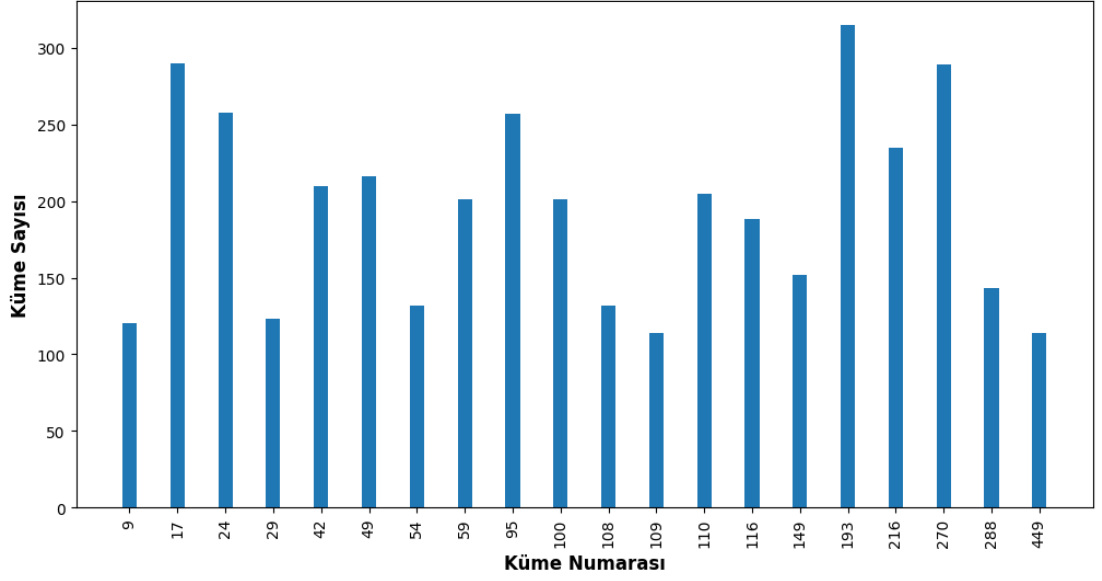
**Tablo 6. Tüm verilerdeki kümelenemeyen yorum sayıları**

Sıra	EPS Deęeri	Küme Sayısı
1	0,6	165379
2	0,7	148793
3	0,8	127766
4	0,9	102130
5	1	21453
6	1,1	0
7	1,2	0
8	1,3	0
9	1,4	0
10	1,5	0
11	1,6	0

Tüm veri setinden yapılan kümeleme çalışmasında EPS değeri 0.6'ya göre en fazla yorum içeren kümeler Tablo 7'de gösterilmektedir.

**Tablo 7. EPS 0.6'ya göre en fazla yorum içeren 20 küme**

Sıra	Küme Numarası	Yorum Sayısı
1	9	120
2	17	290
3	24	258
4	29	123
5	42	210
6	49	216
7	54	132
8	59	201
9	95	257
10	100	201
11	108	132
12	109	114
13	110	205
14	116	188
15	149	152
16	193	315
17	216	235
18	270	289
19	288	143
20	449	114



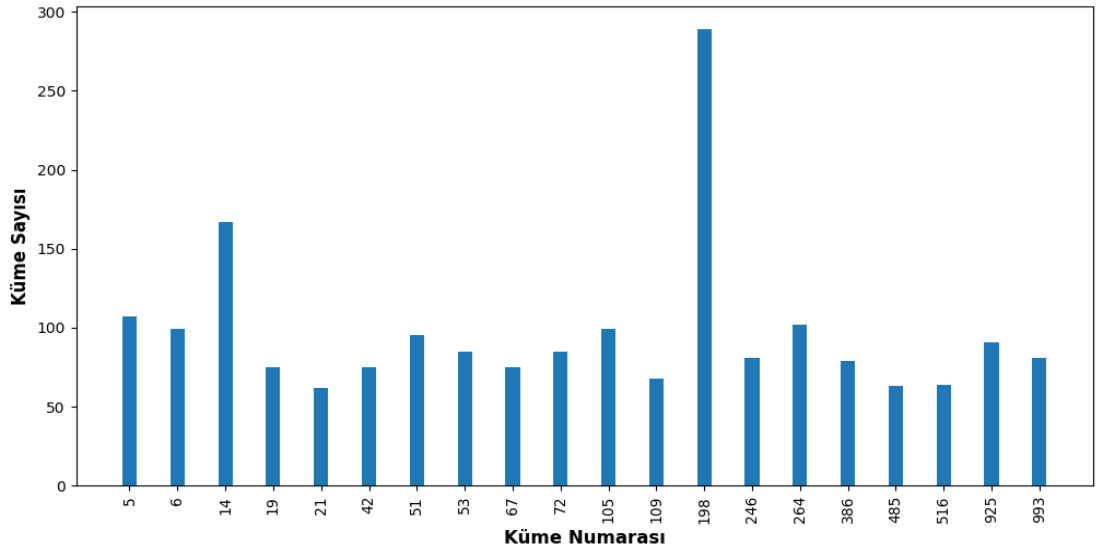
**Şekil 17. EPS 0.6'ya göre en fazla yorum içeren 20 küme**

Tüm veri setinden yapılan kümeleme çalışmasında EPS değeri 0.7'ye göre en fazla yorum içeren kümeler Tablo 8'de gösterilmektedir.

**Tablo 8. EPS 0.7'ye göre en fazla yorum içeren 20 küme**

Sıra	Küme Numarası	Yorum Sayısı
1	5	107
2	6	99
3	14	167
4	19	75
5	21	62
6	42	75
7	51	95
8	53	85
9	67	75
10	72	85
11	105	99
12	109	68
13	198	289

14	246	81
15	264	102
16	386	79
17	485	63
18	516	64
19	925	91
20	993	81



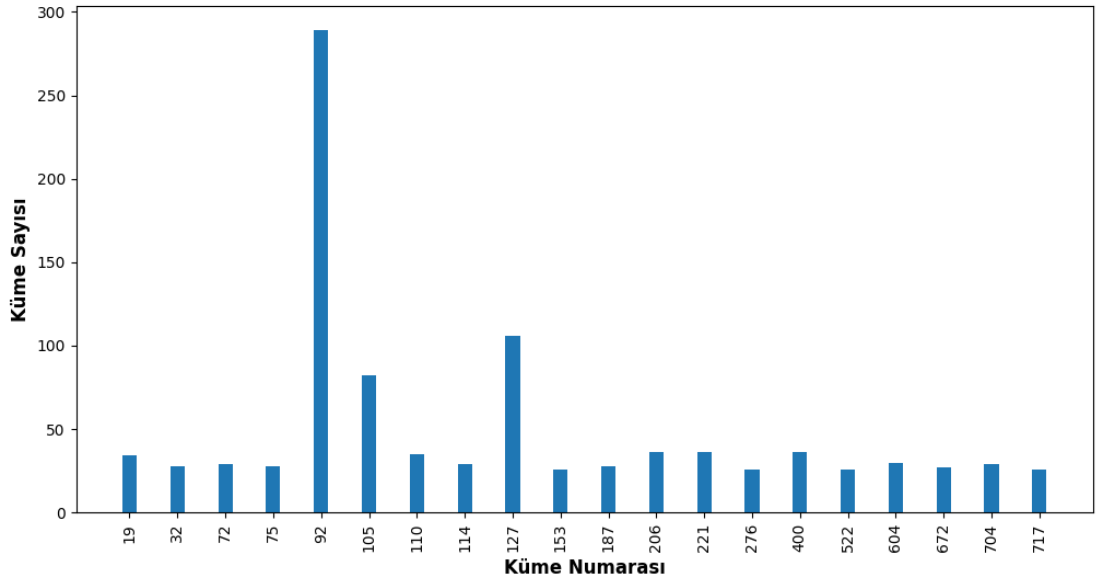
**Şekil 18. EPS 0.7'ye göre en fazla yorum içeren 20 küme**

Tüm veri setinden yapılan kümeleme çalışmasında EPS değeri 0.8'ye göre en fazla yorum içeren kümeler Tablo 9'da gösterilmektedir.

**Tablo 9. EPS 0.8'e göre en fazla yorum içeren 20 küme**

Sıra	Küme Numarası	Yorum Sayısı
1	19	34
2	32	28
3	72	29
4	75	28
5	92	289

6	105	82
7	110	35
8	114	29
9	127	106
10	153	26
11	187	28
12	206	36
13	221	36
14	276	26
15	400	36
16	522	26
17	604	30
18	672	27
19	704	29
20	717	26



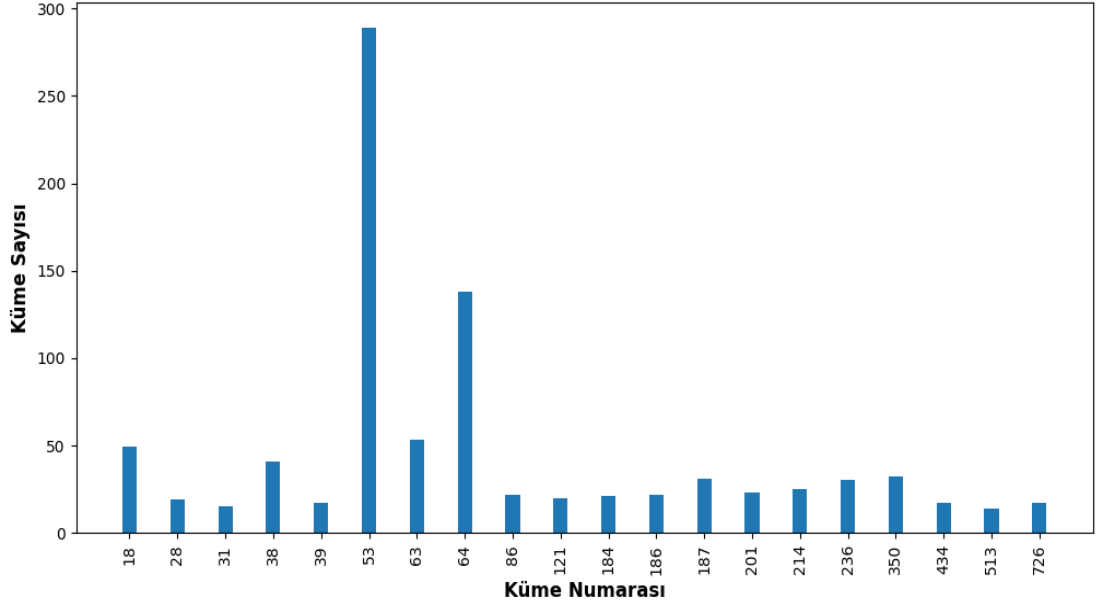
**Şekil 19. EPS 0.8'e göre en fazla yorum içeren 20 küme**

Tüm veri setinden yapılan kümeleme çalışmasında EPS değeri 0.9'a göre en fazla yorum içeren kümeler Tablo 10'da gösterilmektedir.

**Tablo 10. EPS 0.9'a göre en fazla yorum içeren 20 küme**

<b>Sıra</b>	<b>Küme Numarası</b>	<b>Yorum Sayısı</b>
1	18	49
2	28	19
3	31	15
4	38	41
5	39	17
6	53	289
7	63	53
8	64	138
9	86	22
10	121	20
11	184	21
12	186	22
13	187	31
14	201	23
15	214	25
16	236	30
17	350	32
18	434	17
19	513	14
20	726	17



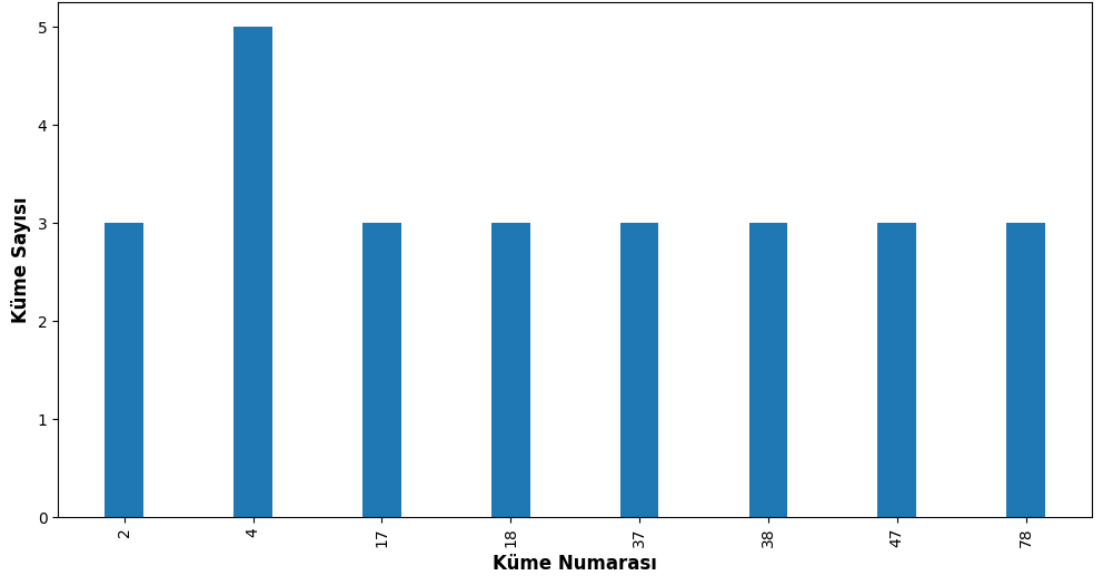


**Şekil 20. EPS 0.9'e göre en fazla yorum içeren 20 küme**

Tüm veri setinden yapılan kümeleme çalışmasında EPS değeri 1.0'e göre en fazla yorum içeren kümeler Tablo 11'te gösterilmektedir.

**Tablo 11. EPS 1.0'e göre en fazla yorum içeren 8 küme**

Sıra	Küme Numarası	Yorum Sayısı
1	2	3
2	4	5
3	17	3
4	18	3
5	37	3
6	38	3
7	47	3
8	78	3



**Şekil 21. EPS 1.0'e göre en fazla yorum içeren 8 küme**

Tüm veri seti ile yapılan kümeleme çalışmalarını detaylı bir şekilde incelemek için yorum örnekleri değerlendirilmiştir. Tablo 12 incelendiğinde yorumlar birebir aynı olduğu ancak puanın farklı olduğu gözlenmektedir. Kümeleme çalışması yorumlara göre yapıldığından küme numaraları aynıdır.

**Tablo 12. Kümeleme örnekleri**

Sıra	Otel Hash	Tarih	Puan	Yorum	EPS 0.6	EPS 0.7	EPS 0.8	EPS 0.9	EPS 1.0
1	433FF	Mayıs 2019	8,0	Otelin konumu, manzarası ve sahili muhteşemdi. Yataklar, çarşafklar, havlular her şey yepyeni ve pırıl pırıldı. Beklentimin çok üstünde mükemmel bir hizmet aldım. Herkese gönül rahatlığıyla tavsiye ederim.	26486	23456	19991	15614	2883
2	433FF	Mayıs 2019	9,3	Otelin konumu, manzarası ve sahili muhteşemdi. Yataklar, çarşafklar, havlular her şey yepyeni ve pırıl pırıldı. Beklentimin çok üstünde mükemmel bir hizmet aldım. Herkese gönül rahatlığıyla tavsiye ederim.	26486	23456	19991	15614	2883

Tablo 13 incelendiğinde yorumların birbirine yakın olduğu ve tarihlerin farklı olduğu gözlenmektedir. Yapılan yorumlar aynı otelde farklı tarih ve aynı puan değerlerine sahiptir.

**Tablo 13. Kümeleme örnekleri**

Sıra	Otel Hash	Tarih	Puan	Yorum	EPS 0.6	EPS 0.7	EPS 0.8	EPS 0.9	EPS 1.0
1	3EB83	Mart 2018	10	Oteli tavsiye ederim. Seda hanım ve ulaş beye iyi niyetleri ve yardımcı olduklarından dolayı teşekkür ediyorum.	42885	38411	33031	26228	5147
2	3EB83	Şubat 2018	10	Oteli tavsiye ediyorum yardımları ve hoşgörüsü için seda hanım ve ulaş beye teşekkür ederim.	42885	38411	33031	26228	5147

Tablo 14 incelendiğinde yorumların birebir aynı olduğu gözlenmektedir. Yapılan yorumlar aynı otelde farklı tarih ve puan değerlerine sahiptir.

**Tablo 14. Kümeleme örnekleri**

Sıra	Otel Hash	Tarih	Puan	Yorum	EPS 0.6	EPS 0.7	EPS 0.8	EPS 0.9	EPS 1.0
1	9C4E4	Ocak 2020	9,7	Büyükada'nın en güzel otellerinden bir tanesi. Temiz, ferah, deniz manzaralı odaları, güler yüzlü çalışanlarıyla hizmette samimi bir otel. Otel işletmecisi fatih bey göstermiş ilgisi gerçekten takdir edilmeye değerdi. Çok teşekkürler yalı butik otel.	4090	3248	2480	1664	71
2	9C4E4	Ekim 2019	9,4	Büyükada'nın en güzel otellerinden bir tanesi. Temiz, ferah, deniz manzaralı odaları, güler yüzlü çalışanlarıyla hizmette samimi bir otel. Otel işletmecisi Fatih Bey göstermiş ilgisi gerçekten takdir edilmeye değerdi. Çok teşekkürler Yalı Butik Otel.	4090	3248	2480	1664	71

Tablo 15 incelendiğinde üç yorumunda aynı otelde yapıldığı gözlenmektedir. Yorumlar birbirine yakın olmasından dolayı aynı küme numaralarına sahiptir.

**Tablo 15. Kümeleme örnekleri**

Sıra	Otel Hash	Tarih	Puan	Yorum	EPS 0.6	EPS 0.7	EPS 0.8	EPS 0.9	EPS 1.0
1	8C438	Aralık 2017	10	Doğru seçim. Otel konum olarak ve kayak pistleri olarak çok güzel her yerde liftler var güzel bir kayak pisti inşa etmişler genel olarak otel kaliteside güzel memnunuz.	2768	2183	1642	1081	47
2	8C438	Mart 2017	10	Doğru seçim. otel konum olarak ve kayak pistleri olarak çok güzel her yerde liftler var güzel bir kayak pisti inşa etmişler genel olarak otel kaliteside güzel memnunuz.	2768	2183	1642	1081	47
3	8C438	Şubat 2017	10	Otel konum olarak ve kayak pistleri olarak çok güzel her yerde liftler var güzel bir kayak pisti inşa etmişler genel olarak otel kaliteside güzel memnunuz.	2768	2183	1642	1081	47

Tablo 16 incelendiğinde iki yorumunda farklı otelde yapıldığı gözlenmektedir. Yorumlar birbirine yakın olmasından dolayı aynı küme numaralarına sahiptir.

**Tablo 16. Kümeleme örnekleri**

Sıra	Otel Hash	Tarih	Puan	Yorum	EPS 0.6	EPS 0.7	EPS 0.8	EPS 0.9	EPS 1.0
1	8C438	Aralık 2017	10	Güler yüzlü ve çok ilgili personeller, tertemiz bir oda, enfes yemekler ve pistlere ulaşım kolaylığı. Herşey çok güzeldi. Teşekkür ederiz kaya ailesinde çok huzurlu bir tatil geçirdik. Önümüzdeki kışa görüşmek üzere.	2784	2198	1656	1095	51
2	5E752	Ocak 2017	10	Harika bir dekorasyon, güler yüzlü ve çok ilgili personeller, tertemiz bir oda, enfes yemekler ve pistlere ulaşım kolaylığı.. Herşey çok güzeldi. Teşekkür ederiz magna ailesi çok huzurlu bir tatil geçirdik. Önümüzdeki kışa görüşmek üzere	2784	2198	1656	1095	51

Tablo 17 incelendiğinde üç yorumunda aynı puana sahip olduğu gözlenmektedir. İlk iki yorum birebir aynı olmasına rağmen farklı otellere aittir. Bu yüzden aynı küme gruplarında toplanmıştır. Üçüncü yorumun bir kısmı diğer yorumlara benzediğinden EPS 1.0 değerine göre aynı küme numaralarına sahiptir.

**Tablo 17. Kümeleme örnekleri**

Sıra	Otel Hash	Tarih	Puan	Yorum	EPS 0.6	EPS 0.7	EPS 0.8	EPS 0.9	EPS 1.0
1	361A9	Ekim 2014	10	Personelin hizmetin çok iyi olmasından bahsederim. Gidilebilecek rahat edilebilecek memnun kalınacak bir tesis olarak tavsiye ederim.	844	698	473	327	18
2	00C5D	Mayıs 2014	10	Personelin hizmetin çok iyi olmasından bahsederim. Gidilebilecek rahat edilebilecek memnun kalınacak bir tesis olarak tavsiye ederim.	844	698	473	327	18
3	361A9	Ekim 2014	10	Termal ve dinlenmek için geldik. Personelin hizmeti çok iyi ve güler yüzlüler. Gidilebilecek rahat edilebilecek memnun kalınacak bir tesis,imkanları oldukça fazla tavsiye ederim.	-1	-1	-1	-1	18

Tablo 18 incelendiğinde iki yorumunda farklı otelde yapıldığı gözlenmektedir. Yorumlar birbirinin aynı olduğu gibi, tarih ve puan değerleri de birebir aynıdır. Yorumlar birebir aynı olmasından dolayı aynı küme numaralarına sahiptir.

**Tablo 18. Kümeleme örnekleri**

Sıra	Otel Hash	Tarih	Puan	Yorum	EPS 0.6	EPS 0.7	EPS 0.8	EPS 0.9	EPS 1.0
1	361A9	Ekim 2016	10	Personel çok güler yüzlüydü misafirperverdi. Hizmeti içten yapıyorlardı. Samimi insanlardı. Resepsiyonda çok güler yüzle karşılandık. Temizliği ve konumundan da çok memnun kaldık.	783	627	410	263	8
2	0B525	Eylül 2016	10	Personel çok güler yüzlüydü misafirperverdi. Hizmeti içten yapıyorlardı. Samimi insanlardı. Resepsiyonda çok güler yüzle karşılandık. Temizliği ve konumundan da çok memnun kaldık.	783	627	410	263	8

Tablo 19 incelendiğinde, iki yorumun da farklı otelde yapıldığı gözlenmektedir. Yorumlar birbirinin aynı olduğu gibi tarih ve puan değerleri de birebir aynıdır. Yorumlar birebir aynı olmasından dolayı aynı küme numaralarına sahiptir.

**Tablo 19. Kümeleme örnekleri**

Sıra	Otel Hash	Tarih	Puan	Yorum	EPS 0.6	EPS 0.7	EPS 0.8	EPS 0.9	EPS 1.0
1	361A9	Temmuz 2016	10	Alışılmıřın dıřında bir konsepti olan bir tesis. Personelin güler yüzlü olması bize özel bir hizmet aldığımızı hissettiriyordu. Yediğimiz ve içtiğimiz her bir kalem malzemede kalite ön plana çıkıyor. Sunum olarak özel bir misafir olduğumuzu hissettirdiler. Temiz bir tesisti. Konum olarak iyiydi.	816	667	450	299	11
2	00C5D	Haziran 2016	10	Alışılmıřın dıřında bir konsepti olan bir tesis. Personelin güler yüzlü olması bize özel bir hizmet aldığımızı hissettiriyordu. Yediğimiz ve içtiğimiz her bir kalem malzemede kalite ön plana çıkıyor. Sunum olarak özel bir misafir olduğumuzu hissettirdiler. Temiz bir tesisti. Konum olarak iyiydi.	816	667	450	299	11

Tablo 20 incelendiğinde, listelenen yirmi üç yorumun farklı otelde yapıldığı gözlenmektedir. Yorumlar birebir aynı olmasından dolayı aynı küme numaralarına sahiptir. Ayrıca EPS 1.0'e göre 0 nolu kümede birbirine benzemeyen çok fazla yorum kümelendiği için 0 nolu küme değerlendirmeye alınmayacaktır. Bu durumda EPS 0.6 ile EPS 0.9 arasında yorumlar üzerinde yapılan kümeleme çalışmasında küme numaralarının aynı çıktığı gözlenmektedir.

**Tablo 20. Kümeleme örnekleri**

Sıra	Otel Hash	Tarih	Puan	Yorum	EPS 0.6	EPS 0.7	EPS 0.8	EPS 0.9	EPS 1.0
1	6F211	Ocak 2007	10	Odalarda kaliteli malzeme vardı.	601	477	301	201	0
2	6F211	Ocak 2007	8,0	Odalarda kaliteli malzeme vardı.	601	477	301	201	0
3	B583D	Ocak 2008	10	Odalarda kaliteli malzeme vardı.	601	477	301	201	0
4	B583D	Nisan 2007	10	Odalarda kaliteli malzeme vardı.	601	477	301	201	0
5	B583D	Ocak 2007	10	Odalarda kaliteli malzeme vardı.	601	477	301	201	0

6	4CCF3	Ocak 2008	10	Odalarda kaliteli malzeme vardı.	601	477	301	201	0
7	1829C	Nisan 2007	7,0	Odalarda kaliteli malzeme vardı.	601	477	301	201	0
8	1084F	Ocak 2007	10	Odalarda kaliteli malzeme vardı.	601	477	301	201	0
9	B85F8	Şubat 2008	7,3	Odalarda kaliteli malzeme vardı.	601	477	301	201	0
10	B85F8	Ocak 2007	10	Odalarda kaliteli malzeme vardı.	601	477	301	201	0
11	7F6A1	Şubat 2008	8,7	Odalarda kaliteli malzeme vardı.	601	477	301	201	0
12	7F6A1	Şubat 2008	6,7	Odalarda kaliteli malzeme vardı.	601	477	301	201	0
13	7F6A1	Ocak 2007	10	Odalarda kaliteli malzeme vardı.	601	477	301	201	0
14	1D149	Şubat 2008	6,0	Odalarda kaliteli malzeme vardı.	601	477	301	201	0
15	1D149	Ocak 2008	9,0	Odalarda kaliteli malzeme vardı.	601	477	301	201	0
16	CA1A0	Ocak 2008	10	Odalarda kaliteli malzeme vardı.	601	477	301	201	0
17	4BA9B	Şubat 2008	8,0	Odalarda kaliteli malzeme vardı.	601	477	301	201	0
18	3B8F3	Ocak 2008	10	Odalarda kaliteli malzeme vardı.	601	477	301	201	0
19	7DC07	Nisan 2007	7,0	Odalarda kaliteli malzeme vardı.	601	477	301	201	0
20	C57F3	Şubat 2008	8,0	Odalarda kaliteli malzeme vardı.	601	477	301	201	0
21	C57F3	Ocak 2008	10	Odalarda kaliteli malzeme vardı.	601	477	301	201	0
22	2F740	Nisan 2007	8,0	Odalarda kaliteli malzeme vardı.	601	477	301	201	0
23	2F740	Ocak 2007	8,0	Odalarda kaliteli malzeme vardı.	601	477	301	201	0

Tablo 21 incelendiğinde, listelenen 21 yorumun farklı otelde yapıldığı gözlenmektedir. Yorumların içeriğine bakıldığında EPS 0.9'a göre "rüya gibiydi" sözcük grubu

kümelemede etkili olduğu gözlenmektedir. Tabloda yorumların farklı olduğu ve diğer EPS değerlerine göre farklı küme numaralarına sahip oldukları görülmektedir.

**Tablo 21. Kümeleme örnekleri**

Sıra	Otel Hash	Tarih	Puan	Yorum	EPS 0.6	EPS 0.7	EPS 0.8	EPS 0.9	EPS 1.0
1	B583D	Şubat 2007	10	Balayı çiftiydik rüya gibiydi.	937	784	544	184	0
2	1829C	Şubat 2007	10	Balayı çiftiydik rüya gibiydi.	937	784	544	184	0
3	B85F8	Şubat 2007	8,0	Balayı çiftiydik rüya gibiydi.	937	784	544	184	0
4	7F6A1	Şubat 2007	10	Balayı çiftiydik rüya gibiydi.	937	784	544	184	0
5	1D149	Şubat 2007	8,0	Balayı çiftiydik rüya gibiydi.	937	784	544	184	0
6	1D149	Şubat 2007	10	Balayı çiftiydik rüya gibiydi.	937	784	544	184	0
7	4BA9B	Şubat 2007	8,0	Balayı çiftiydik rüya gibiydi.	937	784	544	184	0
8	2F740	Ekim 2007	10	Balayı çiftiydik rüya gibiydi.	937	784	544	184	0
9	D626D	Temmuz 2019	10	Çok güler yüzlü hizmet var kaliteli bir otel tam bir butik otel her şey harikaydı rüya gibi bir tatildi.	35112	31408	26991	184	0
10	0B525	Mayıs 2012	10	Harikaydı. Rüya gibiydi. Çalışanları çok çok kibar ve içtendi.	21315	18637	15668	184	0
11	16DD0	Ağustos 2018	10	Hizmet kalite olarak çok güzeldi. Eğlencesi iyiydi. Keyifli bir geziydi. Rüya gibi bir tatildi.	-1	-1	-1	184	0
12	B85F8	Ocak 2007	9,0	Rüya gibi bir oda idi.	2981	2373	1813	184	0
13	5E280	Ocak 2007	10	Rüya gibi bir oda idi.	2981	2373	1813	184	0
14	6F211	Ocak 2008	10	Rüya gibi bir tatildi.	521	416	265	184	0
15	CA1A0	Ocak 2008	8,0	Rüya gibi bir tatildi.	521	416	265	184	0



16	CA1A0	Ocak 2008	9,0	Rüya gibi bir tatildi.	521	416	265	184	0
17	4BA9B	Ocak 2008	8,0	Rüya gibi bir tatildi.	521	416	265	184	0
18	5E280	Ocak 2008	10	Rüya gibi bir tatildi.	521	416	265	184	0
19	3813B	Nisan 2019	10	Rüya gibi bir tatildi.	521	416	265	184	0
20	0B525	Nisan 2019	10	Rüya gibiydi diyebiliriz, genel temizlik ve personel iyiydi.	15604	13361	544	184	0
21	C15BF	Mayıs 2017	10	Rüya gibiydi.	937	784	544	184	0

Tablo 22 incelendiğinde, listelenen yirmi yorumun farklı otele yapıldığı gözlenmektedir. Yorumların içeriğine bakıldığında EPS 0.9'a göre "Kadın erkek havuzlarının ayrı olması" sözcük grubu kümelemede etkili olduğu gözlenmektedir. Tabloda listelenen yorumların farklı olduğu ve diğer EPS değerlerine göre farklı küme numaralarına sahip olduğu görülmektedir. Ayrıca küme numarası -1 olan yorumların, ilgili EPS değerlerine göre kümeleme dışı kaldığı gözlenmektedir.

**Tablo 22. Kümeleme örnekleri**

Sıra	Otel Hash	Tarih	Puan	Yorum	EPS 0.6	EPS 0.7	EPS 0.8	EPS 0.9	EPS 1.0
1	40C1C	Ocak 2018	7,3	Bayan ve erkek havuzlarının ayrı olması çok iyi. Temiz bir mekan.	44213	39596	34042	121	0
2	A725C	Mayıs 2014	9,5	Çocuklar için çok iyi. Havuzu güzel bayan erkek ayrı olması iyi. Genel olarak memnunuz.	-1	-1	174	121	0
3	361A9	Şubat 2018	8,5	Çocuklar için iyiydi tesis. Havuz güzeldi. Kadın erkek ayrı olması güzeldi.	758	603	346	121	0
4	1A421	Haziran 2019	9,0	Fiyat kalite dengesi olarak iyi. Havuz çeşitliliği bayan erkek olarak ayrı olması güzel. Temiz bir otel.	6216	5047	3970	121	0
5	1A421	Ocak 2019	10	Güler yüzlü personelleri ve yardımseverler. Kadın erkek havuzlarının ayrı olması bence iyiydi daha rahat bir ortam oluştu.	6248	5076	3997	121	0

6	361A9	Temmuz 2015	8,0	Hatta bir misafirimi daha göndereceğim. Bayan ve erkek havuzlarının ayrı olması iyiydi. O bölgedeki en iyi tesis burası. Herkese uygun bir tesisti. Temiz bir tesisti.	-1	-1	-1	121	0
7	B583D	Kasım 2019	8,5	Havuzların fazlası olması ve kadın erkek ayrı havuzların olması iyiydi. Temiz bir tesisti. Personeller ilgiliydi. Konum olarak güzeldi.	-1	-1	-1	121	0
8	1A421	Ocak 2020	9,5	Havuzlarının çok olması çok iyiydi. Erkek kadın havuzlarının ayrı olması çok güzeldi. Hamamları çok iyiydi.	6108	4954	3892	121	0
9	361A9	Mart 2018	9,5	Kadın erkek ayrı havuz olması iyiydi. Konumu iyiydi. Personelleri ilgiliydi.	758	603	346	121	0
10	33996	Şubat 2018	9,0	Kadın erkek ayrı olması iyi otelde. Kadın havuzlarına önem vermişler. Genel temizlik iyiydi.	-1	-1	346	121	0
11	361A9	Aralık 2017	9,5	Kadın ve erkek havuzunun ayrı olması güzeldi. Temiz bir tesisti. Personel ilgiliydi.	-1	-1	346	121	0
12	361A9	Ocak 2018	9,0	Konforlu bir otel temizlik olarak beğendim bayan erkek ayrı olması da çok iyi bir şey, genel olarak memnun kaldım.	-1	-1	-1	121	0
13	7F8B9	Şubat 2018	7,5	Konforlu ve rahat bir tesis. Tesiste kadın ve erkek bölümünün ayrı olması güzel. Seçenekli bir tesis. Personel ilgiliydi.	56578	50590	43295	121	0
14	36FB4	Aralık 2016	10	Merkeze yakın bir oteldi. Personelleri iyiydi. Temizliği güzeldi. Havuzların erkek ve kadın ayrı olması güzel değildi. Ortak kullanılamıyordu.	47369	42396	36420	121	0
15	3095F	Ocak 2020	8,0	Odalarda havuz olması iyiydi güzeldi. Bayan erkek ayrı ayrı havuzların olması iyiydi. Personel yaklaşımı gayet güzeldi.	10436	8714	7066	121	0
16	39C05	Aralık 2017	10	Personelleri ilgiliydi. Temiz bir tesisti. Termal olanaklarının kadın erkek ayrı olması çok avantajlıydı çok beğendim.	-1	-1	-1	121	0

17	581AC	Mart 2018	10	Temiz bir oteldi. Bayan ve erkek havuzu ayrı olması güzeldi. Personeller iyiydi.	-1	-1	174	121	0
18	33996	Ocak 2020	9,5	Termal olanaklarında kadın erkek ayrı olması güzeldi. Merkeze yakın olması iyiydi. Temizliği güzeldi.	-1	-1	-1	121	0
19	D970B	Temmuz 2019	9,0	Tesiste kadın erkek için ayrı havuz olması sofi olanlar için güzel temiz bir tesisti konum olarak güzeldi.	-1	-1	346	121	0
20	59A6C	Haziran 2018	9,5	Yeni clupleri güzeldi. Konum olarak bana uygundu. Bayan tuvaletlerini erkek Erkek tuvaletlerini bayan temizlediğini gördüm.	-1	-1	-1	121	0

Tablo 23 incelendiğinde, listelenen 12 yorumun farklı otele yapıldığı gözlenmektedir. Yorumların içeriğine bakıldığında EPS 0.9'a göre "İlgisiz" sözcüğünün kümelemeye etkili olduğu gözlenmektedir. Tabloda yorumların farklı olduğu ve diğer EPS değerlerine göre farklı küme numaralarına sahip olduğu görünmektedir. Ayrıca küme numarası -1 olan yorumlar ilgili EPS değerlerine göre kümeleme dışı kaldığı gözlenmektedir.

**Tablo 23. Kümeleme örnekleri**

Sıra	Otel Hash	Tarih	Puan	Yorum	EPS 0.6	EPS 0.7	EPS 0.8	EPS 0.9	EPS 1.0
1	77DE0	Eylül 2019	2,0	Çok kötüler. İlgisiz alakasız insan halinden anlamayan personel yetkilisi var. Kesinlikle tavsiye etmiyoruz.	-1	-1	-1	586	0
2	6CD4E	Ocak 2018	3,5	İlgisiz bir otel hizmeti çok iyi değil.	-1	1319	932	586	0
3	12CAD	Mayıs 2019	7,0	İlgisiz oldukları için tavsiye etmem olumlu olarak konumu çok güzel.	-1	-1	-1	586	0
4	AFF55	Ağustos 2018	3,3	İlgisiz olmalarının dışında gayet temiz ve güzel bir odaydı.	-1	-1	-1	586	0
5	1A048	Ağustos 2019	2,0	İlgisiz pahalı kısacası çok kötü.	43864	39281	932	586	0
6	0EB9F	Şubat 2018	4,0	İlgisiz personel.	4502	1319	932	586	0

7	4231A	Nisan 2019	6,0	İlgisiz personeller vardı.	4502	1319	932	586	0
8	7FE66	Mayıs 2017	3,3	İlgisiz personeller ve kahvaltı çok kötü.	-1	1319	932	586	0
9	814AB	Temmuz 2019	2,0	İlgisiz ve alakasız personel , müşteri memnuniyeti kesinlikle önemsenmiyor.	55075	1319	932	586	0
10	81EED	Ağustos 2018	3,3	İlgisiz ve alakasız personel.	12063	1319	932	586	0
11	F836E	Aralık 2019	4,7	İlgisiz,temiz bir otel değildi. Şehir manzarası güzeldi.	12093	10189	932	586	0
12	21C36	Kasım 2018	2,7	Odası kötüydü ve pisti. İlgisiz ve alakasızlardı. Konumu fena değildi.	42645	38194	32841	586	0

Tablo 24 incelendiğinde on yorumunda farklı otelde yapıldığı gözlenmektedir. Ayrıca yorumlar birebir aynı olmasından dolayı küme numaraları aynıdır.

**Tablo 24. Kümeleme örnekleri**

Sıra	Otel Hash	Tarih	Puan	Yorum	EPS 0.6	EPS 0.7	EPS 0.8	EPS 0.9	EPS 1.0
1	6F211	Şubat 2008	9,3	O kadar otel gördüm servisi bu kadar güzel otel görmedim.	513	411	262	182	0
2	1D149	Şubat 2008	8,0	O kadar otel gördüm servisi bu kadar güzel otel görmedim.	513	411	262	182	0
3	CA1A0	Şubat 2008	6,7	O kadar otel gördüm servisi bu kadar güzel otel görmedim.	513	411	262	182	0
4	CA1A0	Şubat 2008	4,7	O kadar otel gördüm servisi bu kadar güzel otel görmedim.	513	411	262	182	0
5	4231A	Şubat 2008	10	O kadar otel gördüm servisi bu kadar güzel otel görmedim.	513	411	262	182	0
6	B85F8	Şubat 2008	10	O kadar otel gördüm servisi bu kadar güzel otel grmedim.	513	411	262	182	0
7	1D149	Şubat 2008	6,0	O kadar otel gördüm servisi bu kadar güzel otel grmedim.	513	411	262	182	0

8	1D149	Şubat 2008	8,0	O kadar otel gördüm servisi bu kadar güzel otel grmedim.	513	411	262	182	0
9	4BA9B	Şubat 2008	10	O kadar otel gördüm servisi bu kadar güzel otel grmedim.	513	411	262	182	0
10	5E280	Şubat 2008	6,0	O kadar otel gördüm servisi bu kadar güzel otel grmedim.	513	411	262	182	0

Tablo 25 incelendiğinde on yorumun farklı otele yapıldığı gözlenmektedir. Yorumların içeriğine bakıldığında EPS 0.9'a göre "beklentimin üstündeydi" sözcük grubunun kümelemede etkili olduğu gözlenmektedir. Tabloda yorumların farklı olduğu ve diğer EPS değerlerine göre farklı küme numaralarına sahip olduğu görülmektedir. Ayrıca küme numarası -1 olan yorumlar, ilgili EPS değerlerine göre kümeleme dışı kaldığı gözlenmektedir.

**Tablo 25. Kümeleme örnekleri**

Sıra	Otel Hash	Tarih	Puan	Yorum	EPS 0.6	EPS 0.7	EPS 0.8	EPS 0.9	EPS 1.0
1	EBBF3	Aralık 2019	8,0	Beklentimin üstündeydi.	4889	1301	917	575	0
2	018B1	Haziran 2019	10	Kahvaltısı beklentimin çok çok üstündeydi oda temizliği gayet iyiydi.	-1	1301	917	575	0
3	BB5B6	Aralık 2018	5,3	Oda temiz değildi, dar ve küçüktü. Kahvaltısını beklentimin üstündeydi güzeldi.	-1	-1	-1	575	0
4	9371C	Ağustos 2018	10	Otel beklentimin oldukça üstündeydi. Hizmet kalitesi çok iyiydi.	35914	1301	917	575	0
5	A67A9	Haziran 2018	9,5	Otel çok iyiydi beklentimin çok üstündeydi, temiz bir tesis çalışanlar ilgiliydi konumu iyiydi deniz çok güzeldi.	28990	25767	917	575	0
6	59A6C	Eylül 2017	10	Otel olanakları beklentimin üstündeydi. Her anlamda memnun kaldık.	-1	1301	917	575	0
7	66670	Aralık 2018	9,0	Personel çok iyiydi davranışları güzeldi beklentimin üstündeydi nezih bir yerdı termal olanakları iyiydi.	-1	-1	-1	575	0

8	16DD0	Eylül 2019	8,5	Personeller ilgiliydi. Beklentimin üstündeydi. Temizlik anlamında iyiydi.	-1	1301	917	575	0
9	16DD0	Ağustos 2019	10	Tavsiye ederim. Konforluymdu rahattı. Uygundu beklentimin üstündeydi.	-1	-1	917	575	0
10	CC80A	Ocak 2019	8,0	Temiz bir tesisti. Konumu güzeldi. Oda beklentimin üstündeydi.	4889	1301	917	575	0

Tablo 26 incelendiğinde, listelenen on yorumun farklı otele yapıldığı gözlenmektedir. Yorumların içeriğine bakıldığında EPS 0.9'a göre "mevsim şartlarına" sözcük grubu kümelemede etkili olduğu gözlenmektedir. Tabloda yorumların farklı olduğu ve diğer EPS değerlerine göre farklı küme numaralarına sahip olduğu görülmektedir. Ayrıca küme numarası -1 olan yorumlar, ilgili EPS değerlerine göre kümeleme dışı kaldığı gözlenmektedir.

**Tablo 26. Kümeleme örnekleri**

Sıra	Otel Hash	Tarih	Puan	Yorum	EPS 0.6	EPS 0.7	EPS 0.8	EPS 0.9	EPS 1.0
1	0A9E3	Aralık 2019	10	Çok merkezi konumda her yere yürüme mesafesindeydi temiz bir tesisti. Personeller ilgililerdi mevsim şartlarına göre sıcaklığı yeterliydi.	51648	46186	39557	14	0
2	F3FEF	Aralık 2019	9,0	Konumu merkeziydi. Otoparkı olmaması kötüydü. Temizliği iyiydi . Mevsim şartlarına göre sıcaklığı iyiydi.	-1	59	28	14	0
3	C01C1	Aralık 2019	10	Odanın genişliği güzeldi. Temiz bir tesisti. Mevsim şartlarına göre sıcaklığı iyiydi.	90	59	28	14	0
4	DA3EF	Aralık 2019	9,0	Personelin ilgi ve alakası çok iyiydi. Konumundan memnun kaldım. Mevsim şartlarına göre otelin sıcaklığı uygun ve güzeldi.	-1	-1	28	14	0
5	35256	Aralık 2019	10	Şehir merkezine yakındı. Temiz bir tesisti. Mevsim şartlarına göre sıcaklığı iyiydi.	90	59	28	14	0
6	14C66	Aralık 2019	8,7	Temiz bir tesisti. Konumu güzeldi. Sıcaklığı mevsim şartlarına göre iyiydi.	90	59	28	14	0

7	9AEB6	Ocak 2020	8,0	Temiz bir tesisti. Personeller baya ilgililerdi. Mevsim şartlarına göre sıcaklığı iyiydi.	90	59	28	14	0
8	2EA3F	Aralık 2019	8,7	Temizliği çok güzeldi. Personeller ilgililerdi. Konumu uygundu. Rahat ve konforlu bir tesisti. Mevsim şartlarına göre sıcaklığı iyiydi.	90	59	28	14	0
9	4988B	Aralık 2019	10	Temizliğini çok beğendim çok temizdi. Odanın konumu güzeldi. Mevsim şartlarına göre otelin sıcaklığı iyiydi.	-1	59	28	14	0
10	7FA0C	Aralık 2019	8,7	Ücreti çok pahalıydı. Mevsim şartlarına göre sıcaklığı iyiydi. Konumu merkeziydi.	-1	59	28	14	0

Tablo 27 incelendiğinde, listelenen yedi yorumun farklı otele yapıldığı gözlenmektedir. Ayrıca yorumlar birebir aynı olmasından dolayı küme numaraları da aynıdır.

**Tablo 27. Kümeleme örnekleri**

Sıra	Otel Hash	Tarih	Puan	Yorum	EPS 0.6	EPS 0.7	EPS 0.8	EPS 0.9	EPS 1.0
1	6F211	Şubat 2007	10	Hergun carsaflar deęiřiyordu cok guzeldi.	597	474	300	200	0
2	B85F8	Şubat 2007	6,0	Hergun carsaflar deęiřiyordu guzeldi.	597	474	300	200	0
3	1D149	Şubat 2008	7,3	Hergun carsaflar deęiřiyordu guzeldi.	597	474	300	200	0
4	1D149	Şubat 2007	9,0	Hergun carsaflar deęiřiyordu cok guzeldi.	597	474	300	200	0
5	1D149	Şubat 2007	6,0	Hergun carsaflar deęiřiyordu guzeldi.	597	474	300	200	0
6	7DC07	Şubat 2008	7,3	Hergun carsaflar deęiřiyordu guzeldi.	597	474	300	200	0
7	3ED3D	Şubat 2007	10	Hergun carsaflar deęiřiyordu cok guzeldi.	597	474	300	200	0

Tablo 28 incelendiğinde, listelenen beř yorumun aynı otele yapıldığı gözlenmektedir. Yorumların içeriğine bakıldığında EPS 0.9'a göre "100 üzerinden 100" sözcük grubu kümelemede etkili olduęu gözlenmektedir. Tabloda yorumların farklı olduęu ve dięer EPS deęerlerine göre farklı küme numaralarına sahip olduęu görölmektedir. Ayrıca küme

numarası -1 olan yorumların ilgili EPS değerlerine göre kümeleme dışı kaldığı gözlenmektedir.

**Tablo 28. Kümeleme örnekleri**

Sıra	Otel Hash	Tarih	Puan	Yorum	EPS 0.6	EPS 0.7	EPS 0.8	EPS 0.9	EPS 1.0
1	361A9	Temmuz 2018	10	Bütün hizmetler 100 üzerinden 100 . Teşekkür ederiz.	747	592	384	249	0
2	361A9	Temmuz 2018	10	Bütün hizmetlerden çok memnun kaldık. Hizmetler 100 üzerinden 100.	747	592	384	249	0
3	361A9	Ağustos 2018	10	Her şey için çok teşekkür ederiz. Bütün hizmetler 100 üzerinden 100. Çok memnun kaldık.	747	592	384	249	0
4	361A9	Temmuz 2018	10	Otelde yemekler 100 üzerinden 100. Sadece odalar daha özenli temizlenebilir. Otel genel olarak iyi.	-1	-1	-1	249	0
5	361A9	Eylül 2018	10	Yemekler,odalar,hizmet hepsi kaliteli. 100 üzerinden 100 puan veriyoruz. Teşekkür ederiz.	-1	-1	384	249	0

Tablo 29 incelendiğinde, listelenen beş yorumun farklı otele yapıldığı gözlenmektedir. Yorumların içeriğine bakıldığında EPS 0.9'a göre "Mecbur kalmadıkça" sözcük grubu kümelemede etkili olduğu gözlenmektedir. Tabloda yorumların farklı olduğu ve diğer EPS değerlerine göre farklı küme numaralarına sahip olduğu görülmektedir. Ayrıca, küme numarası -1 olan yorumların, ilgili EPS değerlerine göre kümeleme dışı kaldığı gözlenmektedir.

**Tablo 29. Kümeleme örnekleri**

Sıra	Otel Hash	Tarih	Puan	Yorum	EPS 0.6	EPS 0.7	EPS 0.8	EPS 0.9	EPS 1.0
1	D95F3	Eylül 2015	5,3	Çok mecbur kalmadıkça konum ve yiyeceklerden dolayı kararsızıdm.	-1	-1	-1	535	0
2	9389C	Eylül 2017	4,7	Mecbur kalmadıkça gitmeyin aman.	3863	3099	2361	535	0
3	B3413	Ocak 2018	2,0	Mecbur kalmadıkça gitmeyin.	3863	3099	2361	535	0



4	A1EE1	Şubat 2019	4,0	Sadece yerin uygun olması güzeldi. Mecbur kalmadıkça etmem.	-1	-1	-1	535	0
5	BCEB3	Aralık 2018	7,3	Tesis vasattı. Mecbur kalmadıkça kalınmazdı. Kahvaltısı iyiydi.	-1	-1	2361	535	0

Tablo 30 incelendiğinde, listelenen beş yorumun farklı otele yapıldığı gözlenmektedir. Yorumların içeriğine bakıldığında EPS 0.9'a göre "tereddüt etmeden" sözcük grubunun kümelemede etkili olduğu gözlenmektedir. Tabloda yorumların farklı olduğu ve diğer EPS değerlerine göre farklı küme numaralarına sahip olduğu görülmektedir. Ayrıca, küme numarası -1 olan yorumların, ilgili EPS değerlerine göre kümeleme dışı kaldığı gözlenmektedir.

**Tablo 30. Kümeleme örnekleri**

Sıra	Otel Hash	Tarih	Puan	Yorum	EPS 0.6	EPS 0.7	EPS 0.8	EPS 0.9	EPS 1.0
1	50D8A	Haziran 2019	10	Hiç tereddüt etmeden kalabilirsiniz.	3621	2893	2188	803	0
2	432AE	Mart 2018	10	Hiç tereddüt etmeden kalabilirsiniz.	3621	2893	2188	803	0
3	B2412	Şubat 2019	10	Kesinlikle tereddüt etmeden konaklanabilecek bir otel.	-1	-1	-1	803	0
4	35D66	Temmuz 2019	10	Tereddüt etmeden gidebileceğiniz bir otel. Her şey çok güzeldi.	-1	-1	-1	803	0
5	59A6C	Haziran 2016	10	Tereddüt etmeden tercih edebilirsiniz.	-1	-1	-1	803	0

Tablo 31 incelendiğinde, listelenen dört yorumun aynı puana sahip olduğu gözlenmektedir. İlk iki yorum birebir aynıdır. Diğer iki yorumda farklılık olduğundan dolayı EPS 06. ile EPS 0.7'de küme dışı kalmıştır. EPS 0.8 ile EPS 0.9'da ilk iki yorum ile aynı küme numarasına sahiptir.

**Tablo 31. Kümeleme örnekleri**

Sıra	Otel Hash	Tarih	Puan	Yorum	EPS 0.6	EPS 0.7	EPS 0.8	EPS 0.9	EPS 1.0
1	8C438	Aralık 2017	10	Çalışanlar gayet başarılı ve güler yüzlü odalar gayet temiz ve konforlu housekeeping çalışanları çok ilgililer yemekler muhteşem çok güzel bir hafta sonu geçirdik herşey için teşekkürler.	2771	2186	1645	1084	0
2	8C438	Ocak 2017	10	Çalışanlar gayet başarılı ve güler yüzlü odalar gayet temiz ve konforlu housekeeping çalışanları çok ilgililer yemekler muhteşem çok güzel bir hafta sonu geçirdik herşey için teşekkürler.	2771	2186	1645	1084	0
3	8C438	Ocak 2017	10	çalışanlar gayet başarılı ve güler yüzlü odalar gayet temiz ve konforlu housekeeping çalışanları çok ilgililer yemekler muhteşem çok güzel bir haftasonu geçirdik herşey için teşekkürler. Bir daha ki seneye görüşürüz.	-1	-1	1690	1084	0
4	8C438	Aralık 2016	10	Çalışanlar gayet başarılı ve güler yüzlü odalar gayet temiz ve konforlu housekeeping çalışanları çok ilgililer yemekler muhteşem çok güzel bir haftasonu geçirdik herşey için teşekkürler.	-1	-1	1690	1084	0

Tablo 32 incelendiğinde, listelenen üç yorumun ikisinin birebir aynı olmasına rağmen farklı otellere ait olduğu görülmektedir. EPS 0.8 ile EPS 0.9’da küme numaraları aynıdır.

**Tablo 32. Kümeleme örnekleri**

Sıra	Otel Hash	Tarih	Puan	Yorum	EPS 0.6	EPS 0.7	EPS 0.8	EPS 0.9	EPS 1.0
1	361A9	Ağustos 2016	10	Yakın bir tesisti istanbula yakındı. Lüks ve konforlu. Yemekleri güzeldi. Personel çok güler yüzlüydü. Sabah kahvaltısı ve akşam yemekleri çok iyiydi. Termal havuzları mükemmeldi.	-1	-1	439	289	0
2	361A9	Ağustos 2016	10	Yakın bir tesisti istanbula yakındı. Lüks ve konforlu. Yemekleri güzeldi. Personel çok güler yüzlüydü. Sabah kahvaltısı ve akşam yemekleri çok iyiydi. Termal havuzu daha büyük yada ikiye bölünmüş olabilirdi.	807	656	439	289	0

3	E0F85	Mayıs 2016	9,5	Yakın bir tesisti istanbula yakındı. Lüks ve konforlu. Yemekleri güzeldi. Personel çok güler yüzlüydü. Sabah kahvaltısı ve akşam yemekleri çok iyiydi. Termal havuzu daha büyük yada ikiye bölünmüş olabilirdi.	807	656	439	289	0
---	-------	------------	-----	---	-----	-----	-----	-----	---

Tablo 33 incelendiğinde, listelenen üç yorumun aynı otele ait olduğu görülmektedir. Yorum içeriğindeki farklılıklardan dolayı ikinci yorum EPS 0.8 ile EPS 0.9 değerlerine göre küme numaraları aynıdır. Birinci ve üçüncü yorumlar aynı küme numaraları aynıdır.

**Tablo 33. Kümeleme örnekleri**

Sıra	Otel Hash	Tarih	Puan	Yorum	EPS 0.6	EPS 0.7	EPS 0.8	EPS 0.9	EPS 1.0
1	A725C	Eylül 2019	10	Herşey çok güzel temiz titiz güler yüzlü hizmet.	363	262	158	103	0
2	A725C	Eylül 2019	10	Herşey harika temiz titiz güler yüzlü bir yer çok memnun kaldık.	-1	-1	158	103	0
3	A725C	Eylül 2019	10	Temiz titiz güler yüzlü hizmet aranan herşey var teşekkürler.	363	262	158	103	0

Tablo 34 incelendiğinde, listelenen üç yorumun farklı otellere ait olduğu görülmektedir. Üçüncü yorum diğer yorumlara göre farklı bir içerik barındırdığından dolayı EPS 0.6 ile EPS 0.8 arasındaki değerlere göre yapılan kümeleme çalışmasında elde edilen küme numaraları diğer yorumlardan farklıdır.

**Tablo 34. Kümeleme örnekleri**

Sıra	Otel Hash	Tarih	Puan	Yorum	EPS 0.6	EPS 0.7	EPS 0.8	EPS 0.9	EPS 1.0
1	F2301	Haziran 2018	5,3	Beklentinizi çok yüksek tutmayınız.	3528	2821	2151	1434	0
2	D58F3	Haziran 2019	4,0	Beklentinizi yüksek tutmayın derim.	3528	2821	2151	1434	0
3	07844	Temmuz 2017	6,0	Beklentinizi yüksek tutmazsanız nispeten güzel. Banyosunu pek beğenmedim.	51199	45781	39220	1434	0

#### 4. SONUÇ

Türkiye’de bulunan otel yorumu barındıran servis sağlayıcılarından toplanan veri ile özgün bir veri seti oluşturulan bu tez çalışmasında, DDİ tekniği ile kümeleme algoritmalarından DBSCAN algoritması kullanılarak müşteri incelemelerine yazı kümeleme işlemi uygulanmıştır. Yapılan çalışma üç farklı bölümde incelenmiştir. İlk aşamada otellere ait incelemeler otel bazında kümelenmiştir. Bazı otellere ait incelemenin az olmasından dolayı kümelemede manipülasyona neden olabilecek incelemeler yeterince gözlenememiştir. Sonrasında kümeleme çalışması veri setini beş parçaya bölerek tüm yorumlar üzerinden tekrar gerçekleştirilmiştir. Otel bazlı kümeleme işlemine göre benzer yorumların biraz daha belirgin olduğu tespit edilmiştir. Ancak veri setinin beş parça için ayrı ayrı kümelenmesi, benzer yorumların tüm veri kümesinde tespit edilmesi için yeterli bir sonuç vermemiştir. Bu eksiği giderebilmek adına, son olarak veri setinin tamamına kümeleme işlemi uygulanmıştır. Son yapılan kümeleme çalışmasında tüm veri setinin kullanılması, benzerliklerin tespit edilmesine olanak sağlamaktadır. Çalışmada kendi içerisinde 11 farklı Epsilon parametresi için kümeleme yapılmıştır. DBSCAN algoritmasının parametresi olan komşuluk sorgularında kullanılacak Epsilon değeri 0.6 ile 1.6 arasında 0.1 artış oranı ile kümeleme yapılmıştır. Çalışmada Epsilon değeri 0.6 ve 0.7 olan kümeleme çalışmalarını diğer Epsilon değerlerine göre benzer olan yorumları daha net ortaya çıkardığı görülmüştür.

İlerleyen aşamalarda, başka kümeleme algoritmaları da kullanılarak farklı kümelenmelerin tespit edilmesi konusu araştırmaya açıktır. Bu tür kümelemeler kullanılarak oluşturulabilecek servisler ile müşterilerin yapılan yorumların benzersizliği ve güvenilirliği konusunda daha iyi fikir edinmesi sağlanabilecektir.

## KAYNAKLAR

- Abro, S., Shaikh, S., Abro, R. A., Soomro, S. F., & Malik, H. M. (2020). *Aspect Based Sentimental Analysis of Hotel Reviews: A Comparative Study*. *Sukkur IBA Journal of Computing and Mathematical Sciences*, 4(1), 11-20.
- Adalı, E. (2012). *Doğal Dil İşleme. Türkiye Bilişim Vakfı Bilgisayar Bilimleri ve Mühendisliği Dergisi*, 5(2).
- Ahmetoğlu, H., & Resul, D. A. Ş. (2020). *Türkçe Otel Yorumlarıyla Eğitilen Kelime Vektörü Modellerinin Duygu Analizi ile İncelenmesi*. *Süleyman Demirel Üniversitesi Fen Bilimleri Enstitüsü Dergisi*, 24(2), 455-463.
- Barbado, R., Araque, O., & Iglesias, C. A. (2019). *A framework for fake review detection in online consumer electronics retailers*. *Information Processing & Management*, 56(4), 1234-1244.
- Bose, A., Munir, A., & Shabani, N. (2017). *A Comparative Quantitative Analysis of Contemporary Big Data Clustering Algorithms for Market Segmentation in Hospitality Industry*. arXiv preprint arXiv:1709.06202.
- Bryan Caplan Marketing. (2020). *Top 60 Online Reviews Statistics For 2020*. Erişim adresi: <https://www.bryancaplan.com/blog-posts/top-60-online-reviews-statistics-2020>.
- Carreón, E. C. A., Nonaka, H., & Hiraoka, T. (2019). *Analysis of Chinese Tourists in Japan by Text Mining of a Hotel Portal Site*. arXiv preprint arXiv:1904.13214.
- Chang, Y. C., Ku, C. H., & Chen, C. H. (2019). *Social media analytics: Extracting and visualizing Hilton hotel ratings and reviews from TripAdvisor*. *International Journal of Information Management*, 48, 263-279.
- Chang, Y. C., Ku, C. H., & Chen, C. H. (2020). *Using deep learning and visual analytics to explore hotel reviews and responses*. *Tourism Management*, 80, 104129.
- Chollampatt, S., Taghipour, K., & Ng, H. T. (2016). *Neural network translation models for grammatical error correction*. arXiv preprint arXiv:1606.00189.
- Dave, K., Lawrence, S., & Pennock, D. M. (2003, May). *Mining the peanut gallery: Opinion extraction and semantic classification of product reviews*. In Proceedings of the 12th international conference on World Wide Web (pp. 519-528). ACM.
- Dey, L., Chakraborty, S., Biswas, A., Bose, B., & Tiwari, S. (2016). *Sentiment analysis of review datasets using naive bayes and k-nn classifier*. arXiv preprint arXiv:1610.09982.
- Ester, M., Kriegel, H. P., Sander, J., & Xu, X. (1996, August). *A density-based algorithm for discovering clusters in large spatial databases with noise*. In Kdd (Vol. 96, No. 34, pp. 226-231).

- Fazzolari, M., Cozza, V., Petrocchi, M., & Spognardi, A. (2017). *A study on text-score disagreement in online reviews*. *Cognitive Computation*, 9(5), 689-701.
- Hlee, S., Lee, H., Koo, C., & Chung, N. (2021). *Fake Reviews or Not: Exploring the relationship between time trend and online restaurant reviews*. *Telematics and Informatics*, 59, 101560.
- Hu, Y. H., Chen, K., & Lee, P. J. (2017). *The effect of user-controllable filters on the prediction of online hotel reviews*. *Information & Management*, 54(6), 728-744.
- Ku, C. H., Chang, Y. C., Wang, Y., Chen, C. H., & Hsiao, S. H. (2019, January). *Artificial Intelligence and Visual Analytics: A Deep-Learning Approach to Analyze Hotel Reviews & Responses*. In Proceedings of the 52nd Hawaii International Conference on System Sciences.
- Lin, Y., Lei, H., Wu, J., & Li, X. (2015). *An empirical study on sentiment classification of chinese review using word embedding*. arXiv preprint arXiv:1511.01665.
- Liu, B., Hu, M., & Cheng, J. (2005, May). *Opinion observer: analyzing and comparing opinions on the web*. In Proceedings of the 14th international conference on World Wide Web (pp. 342-351). ACM.
- Ma, Y., Xiang, Z., Du, Q., & Fan, W. (2018). *Effects of user-provided photos on hotel review helpfulness: An analytical approach with deep learning*. *International Journal of Hospitality Management*, 71, 120-131.
- Morinaga, S., Yamanishi, K., Tateishi, K., & Fukushima, T. (2002, July). *Mining product reputations on the web*. In Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 341-349). ACM.
- N. Hu, I. Bose, N.S. Koh, L. Liu, Manipulation of online reviews: an analysis of ratings, readability, and sentiments, *Decis. Support. Syst.* 52 (3) (2012) 674–684.
- Nilashi, M., Ibrahim, O., Yadegaridehkordi, E., Samad, S., Akbari, E., & Alizadeh, A. (2018). *Travelers decision making using online review in social network sites: A case on TripAdvisor*. *Journal of computational science*, 28, 168-179.
- Raguseo, E., Neirotti, P., & Paolucci, E. (2017). How small hotels can drive value their way in infomediation. *The case of 'Italian hotels vs. OTAs and TripAdvisor'*. *Information & Management*, 54(6), 745-756.
- Xu, X. (2020). How do consumers in the sharing economy value sharing? Evidence from online reviews. *Decision Support Systems*, 128, 113162.
- 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.

## ÖZGEÇMİŞ

### KİŞİSEL BİLGİLER

Adı Soyadı : Recep Ali AY

### EĞİTİM DURUMU

Lisans Öğrenimi : 2016, Necmettin Erbakan Üniversitesi, Ahmet Keleşoğlu Eğitim Fakültesi, Bilgisayar ve Öğretim Teknolojileri Eğitimi

Yüksek Lisans Öğrenimi : 2020, KTO Karatay Üniversitesi, Lisansüstü Eğitim Enstitüsü, Elektrik ve Bilgisayar Mühendisliği


Bildiği Yabancı Diller : İngilizce

### İŞ DENEYİMİ

Projeler : Dağıtık Log Yönetimi, Sanal Eleştirel Düşünme Strateji Eğitimi Yazılımı Projesi, Oyun Dünyam Bilgi Depom,

Çalıştığı Kurumlar : 2017 - 2021, Bilgi İşlem Sorumlusu, Konya Sanayi Odası, 2021 – Devam Ediyor, Yazılım Uzmanı, Konya Şehir Hizmetleri A.Ş.

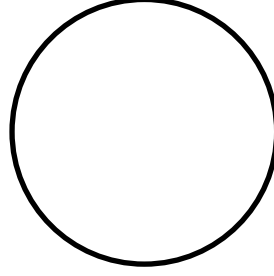
Tarih: 16 Mart 2021

	<p>TEZ/PROJE BAŞLIĞI</p>	<p> KTO Karatay Üniversitesi Lisansüstü Eğitim Enstitüsü ... Anabilim Dalı ... Bilim Dalı</p> <p>TEZ/PROJE BAŞLIĞI</p> <p>Adı SOYADI</p> <p>Yüksek Lisans Tezi/Projesi/Doktora Tezi</p> <p>KONYA Mart 2021</p>
	<p>Adı SOYADI</p>	
	<p>2021</p>	





KTO KARATAY ÜNİVERSİTESİ  
Lisansüstü Eğitim Enstitüsü  
... Anabilim Dalı  
... Bilim Dalı  
Yüksek Lisans / Doktora Programı



Tez Başlığı  
Öğrenci Numarası  
Adı SOYADI  
Danışman: Unvanı Adı SOYADI  
Referans Numarası:



KTO KARATAY ÜNİVERSİTESİ  
Lisansüstü Eğitim Enstitüsü  
... Anabilim Dalı  
... Bilim Dalı  
Yüksek Lisans / Doktora Programı

Tez Başlığı  
Öğrenci Numarası  
Adı SOYADI  
Danışman: Unvanı Adı SOYADI  
Referans Numarası: