**Makine Öğrenmesi ile**

**Seyahat Memnuniyeti Tahmini**

*Bilgisayar Mühendisliği Teknoloji Fakültesi, Gazi Üniversitesi, Ankara, Türkiye mali.cagman@gazi.edu.tr*

*Bilgisayar Mühendisliği, Teknoloji Fakültesi, Gazi Üniversitesi, Ankara, Türkiye icemre.saglam@gazi.edu.tr*

*Bilgisayar Mühendisliği, Teknoloji Fakültesi, Gazi Üniversitesi, Ankara, Türkiye 23181616026@gazi.edu.tr*

**ÖZET**

Günümüz dünyasında seyahat eden bireyler, seyahat ettikleri konumlara dair deneyimlerini dijital platformlar aracılığıyla paylaşarak diğer kullanıcılara geri bildirimde bulundukları için diğer seyahat edecek bireylerin kararlarını da etkileyebilmektedirler hatta etkilemektedirler. Bu makaleye konu olan veri setimizde ise seyahat sektörü ile ilgili kullanıcı deneyimlerinin otomatik olarak analiz edilmesi ve bazı etotlar ile sınıflandırılması amacıyla UCI Machine Learning Repository üzerinde bulunan “Travel Reviews” veri seti kullanılmıştır [1]. Veri setimizde *TripAdvisor* platformunda yer alan 10 farklı kategoriye ait 2048 kullanıcı yorumunu içermektedir [2]. Bu yorumlar, her bir kategoriye ait değerlendirme skorlarıyla birlikte verilmiştir.

İncelenen kategoriler içerisinde sanat galerileri, dans kulüpleri, restoranlar, müzeler, tatil köyleri, parklar, plajlar gibi çeşitli alt başlıklarda geri bildirimler bulunmaktadır. Çalışmamızda ise birbirinden farklı öğrenme algoritmalarının uygulanması sonucunda doğruluk oranları elde edilmiş ve bunların birbirleriyle karşılaştırılması ile en doğru modelin seçilmesi hedeflenmiştir. **Model karşılaştırmalarının sonuçları, her bir algoritmanın performansını gözler önüne sererek, kullanıcı geri bildirimlerini en iyi şekilde sınıflandırabilen algoritmayı belirlememize olanak sağlamıştır.** Bu karşılaştırmalar, **sınıflandırma doğruluğu, algoritmaların genel başarısını ve hangi özelliklerin model üzerinde daha etkili olduğunu** ortaya koymuş, bu sayede seyahat sektörüyle ilgili kullanıcı geri bildirimlerinin daha doğru bir şekilde analiz edilmesine katkıda bulunulmuştur. Kullanılan öğrenme algoritmaları ise Lojistik Regresyon, Karar Ağaçları, Rastgele Orman, KNN, Naive Bayes ve SVM şeklindedir.

Araştırmanın temel amacı, bu kategorilere yönelik kullanıcı yorumlarına dayanarak kullanıcı memnuniyetini sınıflandırmak ve hangi kategorilerin bu memnuniyeti en çok etkilediğini belirlemektir. Bu kapsamda farklı makine öğrenmesi algoritmaları uygulanmış ve en başarılı sınıflandırma modelinin belirlenmesi hedeflenmiştir.

**Anahtar Kelimeler:** Makine Öğrenmesi, Seyahat Hizmetleri, Müşteri Memnuniyeti

**1. Giriş**

Çağımızda, insanların internet siteleri aracılığıyla deneyimlerini paylaşması seyahat ve turizm sektöründe önemli bir bilgi kaynağı olarak kullanılmaktadır [3]. *TripAdvisor* gibi çevrimiçi değerlendirme siteleri kullanıcıların oteller, restoranlar, tarihi eserler gibi seyahat hizmetleri hakkında verdiği puanlar aracılığıyla diğer kullanıcılara yol gösterici niteliğindedir. Bu durumda kullanıcının verdiği puanlar ve yaptığı yorumların analizi; müşteri memnuniyetini ölçme, hizmet kalitesini arttırma, deneyimleyecek olan kullanıcılara kılavuzluk etme açısından kritik bir rol oynamaktadır [4].

Bu çalışmada kullanıcıların *TripAdvisor* platformunda yaptığı değerlendirmeler ele alınarak bir makine öğrenmesi modeli geliştirilmiştir. Bu çalışmanın amacı kullanıcının farklı kategorilerde verdiği puanların yorumlanarak seyahat tercihlerine dair bir öngörü geliştirip bu öngörülerin sınıflama veya kümeleme gibi görevlerde nasıl kullanabileceğini incelemektir. Kullanılan veri seti 10 farklı kategoriye ve bu kategorilere ait puanları içeren çok boyutlu bir yapıya sahiptir [1]. Veri setinin bu yönü sayesinde seyahat deneyimlerinin çeşitli boyutları dikkate alınarak kapsamlı bir analiz yapılabilmektedir. Bu makale kapsamında uygulanan makine öğrenmesi algoritmaları ve ulaşılan sonuçlar ayrıntılı şekilde ele alınacaktır.

**2. Materyal ve Yöntem**

**2.1. Veri Seti**

Dijitalleşmeni hız kazandığı bu dönemde kullanıcıların deneyimlerini çevrimiçi platformlarda paylaşması, turizm ve seyahat gibi sektörlerde veri odaklı analizlerin önünün açılmasını sağlamıştır [4]. Kullanıcıların gezdikleri yerler, kaldıkları oteller, seyahat deneyimleri gibi hizmetler hakkında yaptıkları değerlendirmeler yalnızca kullanıcının bireysel deneyimini yansıtmakla kalmamakta, bu değerlendirmeler sayesinde gelecekteki kullanıcıların kararlarını şekillendirmekte önemli bir referans rolü oynamaktadır [3]. Bu durumda, kullanıcı puanlarının ve yorumlarının sistematik bir şekilde analiz edilmesi, müşteri memnuniyetinin değerlendirilmesi, hizmet verenlerin performanslarının ölçülmesi kritik bir öneme sahiptir.

Çalışmamızda kullanılan veri seti, kullanıcının deneyimlerini paylaştığı *TripAdvisor* platformunda bulunan değerlendirmelere dayanmaktadır. Veri seti toplamda ***2.048 satır*** ve ***10 sütundan*** oluşmaktadır. Her bir satır kullanıcının farklı kategoriler için verdiği puanları temsil ederken; sütunlar sırasıyla şu kategorileri içermektedir:

* Art Galleries
* Dance Clubs
* Juice Bars
* Restaurants
* Museums
* Resorts
* Park/Picnic Spots
* Beaches
* Theaters
* Religious Institutions

Puanlar genelde 0 ile 4 arasında değişmekte olup, kullanıcının bahsi geçen hizmetten ne kadar memnun kaldığını sayısal olarak yansıtmaktadır.

Veri setinin çok boyutlu yapısı sayesinde kullanıcıların seyahat deneyimlerine ilişkin örüntülerini değerlendirmeye uygun bir zemin hazırlamaktadır. Bu çeşitlilik yukarıda bahsi geçen her bir hizmet alanının kullanıcıların memnuniyeti üzerindeki göreli etkisinin karşılaştırılmasına olanak tanımaktadır. Veri seti incelendiğinde kullanıcıların hangi unsurları daha öncelikli gördükleri ve bu önceliklerin kategorilere göre nasıl farklılaştığı fark edilebilmektedir.

Veri seti analiz süreci öncesinde çeşitli ön işleme adımlarından geçirilmiştir. İlk olarak verilerin ölçeklendirilmesi amacıyla **min-max normalizasyon** tekniği kullanılmıştır. Bu yöntemle verilen bütün puanlar 0 ile 1 arasında değerlere dönüştürülerek farklı değişkenler arasında karşılaştırma yapılabilir hale getirilmiştir. Bu adım ile hem veri bütünlüğü korunmuş hem de makine öğrenmesi algoritmasının daha stabil sonuçlar üretmesi sağlanmıştır [5].

Çalışmanın temel amacı doğrultusunda, veri seti üzerinde sınıflandırma yöntemi kullanılarak kullanıcının puanlama davranışlarına göre gruplara ayrılması ve bu gruplar arasındaki benzerliklerin vurgulanması planlanmıştır. Bu analiz sayesinde kullanıcıların profilinin çıkarılması, müşterilerin beklentilerinin daha net anlaşılması ve sektöre yönelik önerilerin geliştirilmesine imkân sağlamıştır. Elde edilen verilerin akademik bir çalışma kapsamında değerlendirilebilmesinin yanı sıra turizm sektörünün profesyonelleri için bir karar destek mekanizması olarak da kullanılabilir niteliktedir.

Bu bağlamda, *TripAdvisor* kullanıcılarından elde edilen bu veri seti; seyahat deneyimlerinin çok boyutlu analizine olanak tanıyan geniş kapsamlı bir içerik sunmakta ve hem verilen hizmetin kalitesi hem de müşterilerin memnuniyeti odaklı çıkarımlar için güçlü bir temel atmaktadır.

Tablo 1. Elde edilen sonuçlar

|  |  |
| --- | --- |
| Değerlendirme Kriterleri | Doğruluk Değerleri |
| K-En Yakın Komşu | 0.9337 |
| Karar Ağacı | **0.9949** |
| Lojistik Regresyon | 0.9490 |
| Rastgele Orman | **0.9949** |
| Destek Vektör Makineleri | 0.9286 |
| Yapay Sinir Ağları | 0.8900 |
| Naive Bayes | 0.8776 |

**2.2. Kullanılan Makine Öğrenmesi Yöntemleri**

Makine öğrenimi (ML), bilgisayarların ve makinelerin insanların öğrenme biçimini taklit etmelerini, görevleri özerk bir şekilde yerine getirmelerini ve deneyim ve daha fazla veriye maruz kalma yoluyla performanslarını ve doğruluklarını artırmalarını sağlamaya odaklanan yapay zekanın bir alt dalıdır [6]. Tipik bir denetlenen makine öğrenmesi algoritması kabaca üç bileşenden oluşur [7]:

1. **Karar süreci:** Verileri alan ve algoritmanızın bulmaya çalıştığı desenin (örüntünün) ne olabileceğini "tahmin eden" bir hesaplama veya adımlar dizisidir.
2. **Hata fonksiyonu:** Tahminin ne kadar iyi olduğunu, bilinen örneklerle karşılaştırarak ölçen bir yöntemdir (bu örnekler mevcutsa). Karar süreci doğru tahmini yaptı mı? Eğer yapmadıysa, "ne kadar kötü" bir tahmin olduğunu nasıl sayısal olarak ifade edebiliriz?
3. **Güncelleme ya da optimizasyon süreci:** Algoritma, yaptığı hatayı inceleyip, bu kararı nasıl verdiğini güncelleyerek bir sonraki seferde hatanın daha az olmasını sağlayan bir yöntemdir.

Makine öğrenmesinin kendine ait farklı şekilde çalışan öğrenme modelleri vardır. Bu çalışmada ise aşağıdaki makine öğrenmesi yöntemleri kullanılmıştır.

1. **Lojistik Regresyon:** Verilen bağımsız değişkenlerle birlikte, bir gözlemin belirli bir sınıfa ait olma olasılığını tahmin etmek için kullanılmıştır. Özellikle ikili sınıflandırma problemlerinde yaygın olarak kullanılan bu yöntem, bu çalışmada çok sınıflı yapıya uyarlanmıştır [8]. Model, veriler arasında doğrusal bir sınır çizerek sınıflandırma işlemi yapar ve olasılıksal çıktılar üretir. Lojistik fonksiyonun (sigmoid) kullanılması sayesinde çıktı değerleri 0 ile 1 arasında normalize edilir.
2. **Karar Ağacı (Decision Tree):** Verileri dallara ayırarak sınıflandıran, yorumlaması kolay bir algoritmadır. Her düğüm bir özelliği, her dal bir kararı ve her yaprak son sınıfı temsil eder. Model, veriyi iteratif olarak bölerek karar kuralları oluşturur ve bu sayede görselleştirilmesi ve anlaşılması kolay bir yapı sunar. Ağaç derinliği ve dallanma sayısı ayarlanarak modelin genelleme kabiliyeti artırılabilir.
3. **K-En Yakın Komşu (KNN):** Sınıflandırma işlemi sırasında, bir veri noktasının en yakın komşularına bakılarak karar verilir. Bu çalışmada farklı k değerleriyle test edilmiştir. KNN, eğitim sırasında model oluşturmaz, tahmin aşamasında tüm eğitim verisini dikkate alır. Bu özelliği sayesinde basit ama veri büyüklüğüne bağlı olarak yavaş çalışabilen bir algoritmadır.
4. **Destek Vektör Makineleri (Support Vector Machine):** Gözetimli öğrenme yöntemlerinden biridir. Genellikle sınıflandırma problemlerinde kullanılır. Bir düzlem üzerine yerleştirilmiş noktaları ayırmak için bir doğru çizer. Bu doğrunun, iki sınıfının noktaları için de maksimum uzaklıkta olmasını amaçlar. Karmaşık ama küçük ve orta ölçekteki veri setleri için uygundur [9].
5. **Rastgele Orman (Random Forest):** Birden fazla karar ağacının birlikte kullanıldığı, güçlü bir topluluk yöntemidir. Karar ağacının aşırı uyum sorununu azaltmak için tercih edilmiştir. Her ağaç rastgele örneklenmiş verilerle ve özelliklerle eğitildiğinden çeşitliliği artırır ve genelleme yeteneğini geliştirir. Son tahmin, tüm ağaçların oy çokluğuna göre belirlenir ve bu da istikrarlı bir performans sağlar.
6. **Naive Bayes:** Özelliklerin birbirinden bağımsız olduğu varsayımıyla çalışan, olasılıksal bir sınıflandırma algoritmasıdır. Basit yapısı sayesinde hızlı sonuç üretmektedir. Bayes teoremini temel alarak, bir gözlemin hangi sınıfa ait olma olasılığının en yüksek olduğunu hesaplar. Özellikle metin sınıflandırma ve spam filtreleme gibi problemler için etkili bir yöntemdir
7. **Min-Max normalizasyonu:** Bu projede, makine öğrenmesi modellerinin daha verimli ve dengeli çalışabilmesi için veri setine uygulanmıştır. Min-Max yöntemiyle tüm özellikler 0 ile 1 aralığına ölçeklendirilmiş ve böylece değişkenler arasındaki farklı büyüklüklerin model üzerindeki etkisi azaltılmıştır [10].
8. **Yapay Sinir Ağları (Artificial Neural Network):** İnsan beyninin çalışma prensibinden esinlenerek geliştirilmiş, çok katmanlı yapay nöronlardan oluşan bir makine öğrenmesi modelidir [11]. Bu projede sınıflandırma problemini çözmek için kullanılmıştır. Modelin aşırı öğrenmesini önlemek amacıyla dropout tekniği uygulanmış ve eğitim sürecinde ‘categorical crossentropy’ kayıp fonksiyonu ile ‘adam’ optimizasyon algoritması kullanılmıştır. Yapay sinir ağı, veriler arasındaki karmaşık ilişkileri öğrenebilme yeteneği sayesinde bu projede anlamlı sonuçlar vermiştir.

Bu çalışmanın sonuçları incelendiğinde hangi algoritmanın hangi doğruluk derecesiyle en iyi sonucu ürettiği bulgular kısmında verilmiştir. Ayrıca gerçekleştirilecek algoritmalar arasında da algoritma başarısına göre de bir sıralama işlemi uygulanacaktır. Makine öğrenmesi yöntemleri ise Python ortamında kodlanmıştır. Python dili, yapay zekâ ve makine öğrenmesi uygulamalarında oldukça popüler ve kullanışlı bir dildir.

**2.3. Literatür Taraması**

Öneri sistemleri, kullanıcıların tercihlerine uygun içerikleri sunarak bilgi aşırı yüklenmesini azaltmayı hedefleyen önemli araçlardır. Bu sistemler, özellikle seyahat sektörü gibi kullanıcı tercihleriyle şekillenen alanlarda büyük önem taşımaktadır.

Adomavicius ve Tuzhilin, öneri sistemlerinin mevcut durumunu ve gelecekteki gelişim alanlarını ele aldıkları çalışmalarında, içerik tabanlı, işbirlikçi ve hibrit yaklaşımların avantajlarını ve sınırlılıklarını detaylı bir şekilde incelemişlerdir. Bu çalışma, öneri sistemlerinin temel yapı taşlarını anlamak için kapsamlı bir kaynak sunmaktadır [12].

Ricci, Rokach ve Shapira tarafından editörlüğü yapılan *Recommender Systems Handbook*, öneri sistemlerinin teorik temelleri, algoritmaları ve uygulama alanları üzerine geniş bir perspektif sunmaktadır. Bu kaynak, özellikle seyahat ve turizm sektöründe öneri sistemlerinin nasıl uygulanabileceğine dair örnekler ve vaka çalışmaları içermektedir [13].

Kurata ve arkadaşları, kullanıcı tarafından oluşturulan içeriklere dayalı olarak kişiselleştirilmiş tur planlaması yapılmasına olanak sağlayan bir yöntem geliştirmişlerdir. Çalışmalarında, önce kullanıcı tercihlerini filtreleyen, ardından bu tercihlere uygun rota oluşturan iki aşamalı bir yapı önerilmiştir. Bu yaklaşım, seyahat planlamasında kullanıcı tercihlerinin nasıl entegre edilebileceğini göstermesi açısından önemlidir [14].

Liu, çok modlu ve çok vektörlü veri madenciliği tekniklerini kullanarak seyahat öneri algoritmaları geliştirmiştir. Bu çalışma, kullanıcıların çeşitli veri kaynaklarından elde edilen bilgilerle daha isabetli öneriler almasını sağlamayı hedeflemektedir [15].

Zhang ve arkadaşları, derin öğrenme tabanlı bir model önererek, kullanıcıların seyahat tercihlerini daha iyi anlamayı ve buna uygun öneriler sunmayı amaçlamışlardır. Bu model, özellikle kullanıcı yorumları ve davranışları üzerinden öğrenme yaparak öneri sistemlerinin doğruluğunu artırmayı hedeflemektedir [16].

Bu çalışmalar, kullanıcı yorumlarına dayalı öneri sistemlerinin geliştirilmesinde farklı yaklaşımların nasıl benimsendiğini ve seyahat sektörü gibi dinamik alanlarda nasıl uygulandığını göstermektedir. Projemiz, bu yaklaşımları temel alarak, kullanıcıların mekân tercihlerine göre öneriler sunmayı ve kullanıcı memnuniyetini artırmayı hedeflemektedir.

**3. Bulgular**

Bu çalışma kapsamında, *TripAdvisor* kullanıcı değerlendirmelerine ait veriler üzerinde çeşitli sınıflandırma modelleri eğitilmiş ve bu modellerin doğruluk oranları karşılaştırılmıştır. Elde edilen sonuçlar doğrultusunda:

* **Karar Ağacı ve Rastgele Orman (Random Forest)** modelleri, sırasıyla %99,49 doğruluk oranı ile en yüksek başarıyı göstermiştir. Bu, her iki modelin de veri setindeki çoklu sınıflandırma problemlerini oldukça başarılı bir şekilde çözdüğünü ortaya koymaktadır. Sosyal hayatta bu modellerin kullanımı, seyahat ve turizm sektöründe kullanıcıların tercihlerini daha doğru bir şekilde tahmin edebilme ve kişiselleştirilmiş öneriler sunma açısından önemli bir avantaj sağlayabilir.
* **Lojistik Regresyon** modeli, %94,90 doğruluk oranı ile yüksek bir performans sergilemiş, ancak daha karmaşık modellerin gerisinde kalmıştır. Bu model, sosyal medya analizlerinde ve kullanıcı geri bildirimlerinin sınıflandırılmasında kullanılabilir, ancak daha gelişmiş algoritmalarla daha doğru tahminler yapılabilir.
* **Destek Vektör Makineleri (SVM)**, %92,86 doğruluk oranı ile başarılı bir sonuç elde etmiş, ancak bu modelin performansı Karar Ağacı ve Rastgele Orman modelleriyle karşılaştırıldığında daha düşük kalmıştır. SVM, daha küçük ve daha homojen veri setlerinde etkili olabilir, ancak büyük ve karmaşık veri setlerinde daha düşük performans sergileyebilir. Bu, sosyal ağlardaki kullanıcı etkileşimlerini sınıflandırmada, ancak sınırlı veri üzerinde oldukça faydalı olabilir.
* **K-En Yakın Komşu (KNN)** algoritması, %93,37 doğruluk oranı ile yerine rastgele orman ve karar ağacına göre daha düşük bir başarı göstermiştir. Sosyal medya platformlarında kullanıcıların benzerliklerine dayalı önerilerde kullanılabilir, ancak verinin yoğunluğuna bağlı olarak performansı değişkenlik gösterebilir.
* **Naive Bayes** modeli, %87,76 doğruluk oranı ile en düşük başarıyı elde etmiştir. Bu model, kullanıcı davranışlarının ve yorumlarının basit kategorilere ayrılmasında etkili olabilir, ancak daha karmaşık sınıflandırma görevlerinde yeterli performansı gösteremeyebilir.
* **Yapay Sinir Ağları (ANN)** modeli, bu çalışmada %89 doğruluk oranı ile orta düzeyde bir performans sergilemiştir. Modelin özellikle veri içerisinde baskın olan sınıfları başarıyla öğrenebildiği, ancak düşük temsil edilen sınıflarda zayıf kaldığı gözlemlenmiştir. ANN, çok katmanlı yapısı sayesinde doğrusal olmayan ilişkileri modelleyebilme kapasitesine sahiptir; bu nedenle daha büyük ve dengeli veri setlerinde daha yüksek başarı potansiyeline sahiptir.

Bu bulgular, özellikle Karar Ağacı ve Rastgele Orman modellerinin çok sınıflı sınıflandırma problemlerinde yüksek doğruluk oranlarıyla etkili sonuçlar elde ettiğini göstermektedir.

Gerçekleştirilen veri analizi kapsamında, kullanıcıların çeşitli mekanlara yönelik ortalama geri bildirim puanları üzerinden oluşturulan değişkenlerin, tahmin modeline olan katkıları incelenmiştir. Random Forest algoritması ile yürütülen özellik önem analizi sonucunda, **Category 10 (Dini kurumlar üzerine kullanıcı geri bildirimleri)** değişkeninin, modelin karar mekanizmasında en yüksek öneme sahip olduğu tespit edilmiştir. Özellik önem değeri yaklaşık **%55** seviyesinde gerçekleşen bu değişkeni, sırasıyla **Category 7 (Parklar ve piknik alanları)** ve **Category 8 (Plajlar)** değişkenleri takip etmektedir. Buna karşın **Category 9 (Tiyatrolar)** ve **Category 2 (Dans kulüpleri)** gibi değişkenlerin model üzerindeki etkisinin oldukça sınırlı olduğu gözlemlenmiştir.

Elde edilen bu bulgu, kullanıcı tercihlerini ve memnuniyet düzeylerini tahmin etme sürecinde, özellikle dini kurumlara yönelik geri bildirimlerin belirleyici bir faktör olduğunu ve modelin bu değişkene diğerlerine kıyasla çok daha fazla ağırlık verdiğini göstermektedir. Bu sonuç, ilgili kategorilerin kullanıcı davranışları üzerindeki etkisini değerlendirmek ve gelecekteki analizlerde önceliklendirme yapmak açısından anlamlı bir gösterge olarak değerlendirilebilir.

Şekil 1. Modellerin Doğruluk Oranlarının Karşılaştırılması

metin, ekran görüntüsü, öykü gelişim çizgisi; kumpas; grafiğini çıkarma, çizgi içeren bir resim

Yapay zeka tarafından oluşturulan içerik yanlış olabilir.

Şekil 2. Ortalama Puanlara Göre Kategoriler

metin, ekran görüntüsü, diyagram, öykü gelişim çizgisi; kumpas; grafiğini çıkarma içeren bir resim

Yapay zeka tarafından oluşturulan içerik yanlış olabilir.

Şekil 3. K Değerine Göre KNN Doğruluk Oranları Şekil 4. Rastgele Orman Modeli Özelliklerinin Önem Dereceleri

metin, ekran görüntüsü, ekran, görüntüleme, çizgi içeren bir resim

Yapay zeka tarafından oluşturulan içerik yanlış olabilir.çizgi, öykü gelişim çizgisi; kumpas; grafiğini çıkarma, diyagram, metin içeren bir resim

Yapay zeka tarafından oluşturulan içerik yanlış olabilir.

Şekil 5. Rastgele Orman Modelinin Karar Ağacından Bir Kesit

diyagram, mor içeren bir resim

Yapay zeka tarafından oluşturulan içerik yanlış olabilir.

Şekil 6. Karar Ağacı Modeli

metin, Post-it notu, ekran görüntüsü, el yazısı içeren bir resim

Yapay zeka tarafından oluşturulan içerik yanlış olabilir.

**4. Sonuç ve Öneriler**

Bu çalışmada *TripAdvisor* platformunun veri seti kullanılarak kullanıcıların seyahat deneyimlerine yönelik değerlendirmeleri analiz edilmiştir. Bu amaçla, altı farklı makine öğrenmesi modeli kullanılarak kullanıcı davranışları sınıflandırılmaya çalışılmıştır. Min-Max normalizasyonu tekniği kullanılarak veriler ölçeklendirilmiş; sınıflandırma yöntemi yardımıyla veri modeli oluşturulmuştur. Analizde kullanılan makine öğrenmesi yöntemleri için 2018 yılına ait *TripAdvisor* platformundaki 980 kullanıcıya ait veriler kullanılmıştır. Bu veriler %80 eğitim ve %20 test olacak şekilde ikiye ayrılarak değerlendirilmiştir. Elde edilen sonuçlara göre Karar Ağacı (%99,49) ve Rastgele Orman (%99,49) yöntemleri en yüksek doğruluk oranına ulaşarak en başarılı modeller olmuştur. Lojistik regresyon (%94,90), SVM (%92,86) ve KNN (%93,37) algoritmaları da yüksek doğruluk oranları ile tatmin edici sonuçlar vermiştir. En düşük performanslar ise %87,76 doğruluk oranı ile Naive Bayes ve %89,00 doğruluk oranı ile Yapay Sinir Ağları yöntemlerinde gözlemlenmiştir. Bu çalışma ile makine öğrenmesi tekniklerinin, seyahat hizmetlerindeki müşteri memnuniyetini analiz etmede ne kadar etkili olduğu vurgulanmıştır. Özellikle karar ağacı ve rastgele orman modelleri, yüksek doğruluk oranları sebebiyle, seyahat hizmetlerindeki müşteri memnuniyetini tahmin etmede güvenilir bir tahmin mekanizması sunmaktadır. Elde edilen sonuçlar hem akademik hem de sektörel olarak yürütülen uygulamalara katkı sağlayabilecek niteliktedir.

**KAYNAKÇA**

* 1. D. Dua and C. Graff, “UCI Machine Learning Repository,” *University of California, Irvine, School of Information and Computer Sciences*, 2019. [Online]. Available: <https://archive.ics.uci.edu/ml/datasets/Travel+Reviews>
  2. TripAdvisor. [Online]. Available: <https://www.tripadvisor.com>
  3. H. Gupta, "Predicting sentiment and rating of tourist reviews using machine learning," *ResearchGate*, 2023. [Online]. Available: <https://www.researchgate.net/publication/361997564_Predicting_sentiment_and_rating_of_tourist_reviews_using_machine_learning>. [Accessed: May 14, 2025].
  4. K. R. S. Kumar, M. N. S. Kumari, and S. R. N. K. Aruna, "Sentiment analysis of online tourist reviews using machine learning," *Journal of Hospitality and Tourism Insights*, vol. 7, no. 1, pp. 1-15, 2022.
  5. G. J. L. Wang and Y. S. Tan, "Data preprocessing in machine learning algorithms: A survey," *International Journal of Computer Applications*, vol. 56, no. 7, pp. 23-32, 2018.
  6. A. Géron, *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow*, 2nd ed., O'Reilly Media, 2019.
  7. T. Mitchell, *Machine Learning*, McGraw-Hill, 1997.
  8. S. Raschka and V. Mirjalili, *Python Machine Learning*, 3rd ed., Packt Publishing, 2019.
  9. C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995.
  10. I. Guyon, G. Elisseeff, “An introduction to variable and feature selection,” *Journal of Machine Learning Research*, vol. 3, pp. 1157–1182, 2003.
  11. F. Chollet, *Deep Learning with Python*, 2nd ed., Manning Publications, 2021.
  12. G. Adomavicius and A. Tuzhilin, “Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions,” \*IEEE Transactions on Knowledge and Data Engineering\*, vol. 17, no. 6, pp. 734–749, Jun. 2005.
  13. F. Ricci, L. Rokach, and B. Shapira, Eds., \*Recommender Systems Handbook\*, 2nd ed., Springer, 2015.
  14. T. Kurata, M. Tsubouchi, H. Shintani, and M. Tsukamoto, “CT-Planner5: Personalized tour planning system using geotagged tweets,” in \*Proc. IEEE/WIC/ACM Int. Conf. Web Intelligence (WI)\*, Oct. 2013, pp. 321–328.
  15. X. Liu, Y. Liu, S. Liu, and H. Xiong, “SmarTraveler: A tourist trip planning system with multi-modal route recommendations,” \*ACM Transactions on Intelligent Systems and Technology\*, vol. 8, no. 2, pp. 1–29, Nov. 2016.
  16. Y. Zhang, G. Lai, M. Zhang, Y. Zhang, Y. Liu, and S. Ma, “Explicit factor models for explainable recommendation based on phrase-level sentiment analysis,” in \*Proc. 37th Int. ACM SIGIR Conf. Research & Development in Information Retrieval\*, 2014, pp. 83–92.