

Paylaşımlı Bisiklet Sayısı Tahmininde Makine Öğrenmesi Tekniklerinin Performans Karşılaştırması

Bilgisayar Mühendisliği, Teknoloji Fakültesi, Gazi Üniversitesi, Ankara, Türkiye
myasin.alarslan@gazi.edu.tr

Bilgisayar Mühendisliği, Teknoloji Fakültesi, Gazi Üniversitesi, Ankara, Türkiye
falperen.kilic@gazi.edu.tr

ÖZET

Kentleşmenin hızla artması, büyük şehirlerde trafik yoğunluğu, hava kirliliği ve fosil yakıt tüketimi gibi ulaşım sorunlarını beraberinde getirmiştir. Bu sorunlara alternatif bir çözüm olarak bisiklet paylaşım sistemleri (BPS), çevre dostu, ekonomik ve sürdürülebilir bir ulaşım seçeneği sunmaktadır. Pek çok kent, bu sistemleri şehir içi ulaşımında tamamlayıcı bir araç olarak benimsemektedir.

Bu çalışmada, Washington D.C.'deki Capital Bikeshare sistemine ait 2011-2012 yıllarını kapsayan ve UCI Machine Learning Repository tarafından sağlanan "Bike Sharing" veri seti kullanılarak bisiklet kiralama talebinin tahmin edilmesi amaçlanmıştır. Veri seti; mevsim, sıcaklık, nem, rüzgar hızı, tatil ve iş günü durumu gibi zamansal ve çevresel değişkenleri içermektedir.

Araştırma kapsamında veri ön işleme adımları uygulanmış, eksik ve aykırı değerler temizlenmiş ve min-max normalizasyonu gerçekleştirilmiştir. Ardından Karar Ağacı, Rastgele Orman, K-En Yakın Komşu (K-NN), Lineer Regresyon ve Naive Bayes olmak üzere beş farklı denetimli makine öğrenmesi algoritması kullanılarak tahmin modelleri oluşturulmuştur. Modellerin başarımı çeşitli değerlendirme metrikleriyle karşılaştırılmıştır.

Elde edilen bulgular, Karar Ağacı ve Rastgele Orman algoritmalarının diğer yöntemlere göre daha yüksek doğruluk sağladığını göstermektedir. Bu sonuçlar, bisiklet paylaşım sistemlerinin talep tahmini ve planlamasında makine öğrenmesi yöntemlerinin etkili bir biçimde kullanılabileceğini ortaya koymaktadır.

Anahtar Kelimeler: Makine Öğrenmesi, Talep Tahmini, Bisiklet Paylaşımı, Regresyon Analizi, UCI Veri Seti

1. Giriş

Kentleşmenin hızla artması, dünya genelinde büyük şehirlerde ulaşım sistemlerine olan baskıyı artırmakta ve beraberinde çevresel, ekonomik ve sosyal sorunları da gündeme getirmektedir. Özellikle trafik yoğunluğu, hava kirliliği ve fosil yakıt tüketimi gibi problemler, sürdürülebilir ulaşım çözümlerine olan ihtiyacı her geçen gün daha da önemli hale getirmektedir[7]. Bu doğrultuda, çevre dostu, düşük maliyetli ve pratik bir ulaşım alternatifi sunan bisiklet paylaşım sistemleri (BPS), son yıllarda pek çok şehirde yaygınlaşmıştır [4]. Bisiklet paylaşım sistemleri, kısa mesafeli yolculuklar için hızlı erişim sağlayarak karbon salımını azaltmakta, ulaşımın çevresel etkilerini hafifletmekte ve toplumsal sağlık düzeyine olumlu katkılarda bulunmaktadır [8].

Ancak bu sistemlerin sürdürülebilir bir şekilde işletilebilmesi, hizmet kalitesinin korunması ve kullanıcı memnuniyetinin sağlanması açısından kullanıcı talebinin doğru bir şekilde tahmin edilmesi kritik bir öneme sahiptir. Günün farklı saatlerinde veya yılın farklı zamanlarında talepte yaşanan dalgalanmalar, istasyonlardaki bisiklet arz-talep dengesini olumsuz etkileyebilmekte; bu da kullanıcıların ihtiyaç duydukları anda bisiklet bulamamaları veya istasyonlara bisiklet bırakamamaları gibi sorunlara yol açabilmektedir [5]. Bu bağlamda, talep tahmini yapabilen sistemlerin geliştirilmesi, kaynakların daha verimli kullanılmasını sağlayarak sistemlerin etkinliğini ve kullanıcı memnuniyetini artıracaktır [15].

Bu çalışmanın temel amacı, bisiklet paylaşım sistemlerinde kullanıcı talebini etkileyen faktörleri belirlemek ve bu faktörlere dayalı olarak talep tahmini gerçekleştirmektir. Araştırma kapsamında şu sorulara yanıt aranmıştır:

- Hangi çevresel ve zamansal değişkenler kullanıcı talebini en çok etkilemektedir?
- Farklı makine öğrenmesi algoritmaları ile talep tahmini ne ölçüde başarılı bir şekilde gerçekleştirilebilir?
- Bu algoritmaların karşılaştırmalı performansı nasıldır ve hangi yöntem öngörü açısından daha güvenilirirdir?

Bu amaçla, UCI Machine Learning Repository üzerinden elde edilen ve Washington D.C. şehrindeki bisiklet paylaşım sistemine ait verileri içeren Bike Sharing Dataset kullanılmıştır [3]. Veri seti; sıcaklık, nem, rüzgar hızı gibi meteorolojik değişkenlerin yanı sıra mevsim, tatil günleri, haftanın günleri ve saat gibi zamansal bileşenleri içermektedir. Çalışmada bu çok boyutlu veri seti üzerinde Karar Ağacı, Rastgele Orman, K-En Yakın Komşu (K-NN), Lineer Regresyon ve Naive Bayes gibi farklı makine öğrenmesi yöntemleri uygulanmış ve bu modellerin kullanıcı talebini tahmin etme başarıları karşılaştırmalı olarak analiz edilmiştir.

Bu çalışmanın önemi, yalnızca talep tahminine katkı sağlamakla sınırlı kalmayıp aynı zamanda şehir içi ulaşım planlamasına, hizmet kalitesinin iyileştirilmesine ve çevre dostu ulaşım politikalarının desteklenmesine yönelik somut girdiler sağlamasında yatmaktadır. Ayrıca, geliştirilen model ve elde edilen bulgular, gelecekte farklı şehirler ve ulaşım sistemlerine uyarlanabilecek veri odaklı karar destek sistemleri için de temel teşkil etmektedir.

2. Materyal ve Yöntem

2.1. Veri Seti

Günümüzde hızla artan kentleşme, dünya genelinde büyük şehirlerdeki ulaşım altyapısı üzerinde ciddi bir baskı oluşturmaktadır. Artan nüfus yoğunluğu, bireysel araç kullanımının yaygınlaşması ve altyapının bu büyümeye yeterince ayak uyduramaması; trafik sıkışıklığı, zaman kayıpları, hava kirliliği, karbon salınımı ve fosil yakıt tüketimi gibi pek çok çevresel ve ekonomik problemi de beraberinde getirmektedir. Özellikle ulaşım kaynaklı karbon salınımı, iklim değişikliğiyle mücadele kapsamında kritik öneme sahip bir sorun haline gelmiştir. Bu

bağlamda, sürdürülebilir ulaşım çözümlerinin geliştirilmesi ve yaygınlaştırılması hem çevresel hem de toplumsal refah açısından giderek daha büyük bir gereklilik halini almaktadır.

Bu ihtiyaç doğrultusunda geliştirilen **bisiklet paylaşım sistemleri**, düşük maliyetli, çevre dostu ve fiziksel hareketliliği destekleyen alternatif bir ulaşım biçimi olarak ön plana çıkmaktadır. Modern şehircilik yaklaşımlarında yaygınlaşan bu sistemler, ulaşımın son kilometresi olarak bilinen "last-mile transportation" sorununa da etkili çözümler sunmaktadır. Ancak bu sistemlerin verimli ve sürdürülebilir şekilde işletilebilmesi için kullanıcı talebinin doğru biçimde tahmin edilmesi büyük önem taşımaktadır. Zira talebin doğru öngörülememesi durumlarda, istasyonlar arasında dengesizlikler, boş ya da aşırı dolu bisiklet noktaları, kullanıcı memnuniyetsizliği ve operasyonel verimsizlik gibi sorunlar ortaya çıkabilmektedir. Bu nedenle, kullanıcı davranışlarını anlamaya yönelik veri odaklı analizler ve tahmin modelleri, bu sistemlerin başarısında belirleyici rol oynamaktadır.

Bu çalışmada, bisiklet paylaşım sistemlerinde kullanıcı talebinin tahmin edilmesine yönelik analizler gerçekleştirmek amacıyla **UCI Machine Learning Repository** üzerinde yer alan ve yaygın olarak kullanılan **Bike Sharing Dataset** adlı veri seti kullanılmıştır (Fanaee-T & Gama, 2014). Veri seti, 2011–2012 yılları arasında Amerika Birleşik Devletleri'nin başkenti Washington D.C.'de hizmet veren **Capital Bikeshare** sisteminden elde edilen günlük verileri içermektedir. İstasyon bazlı çalışan bu sistemde, kullanıcılar herhangi bir istasyondan bisiklet kiralarak farklı bir noktaya bırakabilmekte, böylece kısa mesafeli ulaşım ihtiyaçlarını karşılayabilmektedirler.

Veri seti, toplam **731 günlük gözlem** ve **16 değişken** içermektedir. Kullanıcıların bisiklet kiralama davranışlarını etkileyebilecek çok sayıda faktörü kapsayan bu değişkenler; meteorolojik koşullar (sıcaklık, hissedilen sıcaklık, nem, rüzgar hızı), zamansal bileşenler (mevsim, ay, hafta içi/günü, yıl), sosyal göstergeler (tatil günü, çalışma günü) ve kullanıcı türlerine göre ayrıştırılmış kiralama sayılarından oluşmaktadır. Çalışmada **bağımlı değişken** olarak cnt (toplam günlük kiralama sayısı) seçilmiştir. Bu değişken, casual (üye olmayan kullanıcılar) ve registered (üyelikli kullanıcılar) olmak üzere iki gruba ayrılmış kiralama sayılarını toplamaktadır.

Veri seti, yapısal olarak oldukça düzenlidir ve eksik veri içermemektedir. Ayrıca bazı değişkenler, özellikle sıcaklık ve rüzgar gibi çevresel ölçütler, makine öğrenmesi algoritmalarında daha iyi performans elde edilebilmesi amacıyla **min-max normalizasyonu** yöntemiyle **[0, 1]** aralığında ölçeklendirilmiştir. Bununla birlikte, season, weathersit ve weekday gibi bazı kategorik değişkenlerin sayısal formatta verilmiş olması, modelleme öncesinde bu değişkenlerin uygun dönüşüm yöntemleriyle işlenmesini gerektirmiştir.

Veri Setindeki Değişkenler

Değişken Adı	Açıklama	Veri Türü	Detay
instant	Gözlem numarası (otomatik artan indeks)	Tamsayı (Integer)	Sıralama için
dteday	Tarih bilgisi (YYYY-MM-DD formatında)	Tarih (Date)	Zaman serisi analizi için
season	Mevsim (1: İlkbahar, 2: Yaz, 3: Sonbahar, 4: Kış)	Kategorik (Ordinal)	Mevsimsel farklılıklar
yr	Yıl (0: 2011, 1: 2012)	İkili (Binary)	Zaman etkisi
mnth	Ay (1–12 arası)	Sayısal (Ordinal)	Aylık dönemsel analiz
holiday	Resmi tatil günü mü? (1: Evet, 0: Hayır)	İkili (Binary)	Kullanım deseni değişebilir
weekday	Haftanın günü (0: Pazar – 6: Cumartesi)	Sayısal (Ordinal)	Haftalık kullanım farkları
workingday	Çalışma günü (1: Hafta içi ve tatil olmayan günler)	İkili (Binary)	Yoğunluk analizi için önemli
weathersit	Hava durumu: 1: Açık2: Bulutlu3: Yağışlı4: Fırtınalı	Kategorik (Ordinal)	Hava etkisi önemli
temp	Sıcaklık (0–1 arası, 41’e bölünmüş)	Sürekli (Float)	Fiziksel konfor faktörü
atemp	Hissedilen sıcaklık (0-1 arası, 50’ye bölünmüş)	Sürekli (Float)	Algılanan sıcaklık
hum	Nem oranı (0–1 arası, 100’e bölünmüş)	Sürekli (Float)	Fiziksel etkiler
windspeed	Rüzgar hızı (0-1 arası, 67’ye bölünmüş)	Sürekli (Float)	Konfor koşulları
casual	Üye olmayan kullanıcıların kiralama sayısı	Tamsayı (Integer)	Turistik/tek kullanımlık
registered	Üye olan kullanıcıların kiralama sayısı	Tamsayı (Integer)	Düzenli kullanıcılar
cnt	Günlük toplam kiralama sayısı (casual + registered)	Tamsayı (Integer)	Hedef değişken (output)

Veri seti bu yönleriyle hem regresyon hem de sınıflandırma problemleri için elverişli bir yapıya sahiptir. Farklı makine öğrenmesi modellerinin eğitime uygun olması, geçmişe dayalı kullanıcı davranışlarının analiz edilmesine ve gelecekteki taleplerin daha doğru bir şekilde tahmin edilmesine olanak tanımaktadır.

2.2. Kullanılan Makine Öğrenmesi Algoritmaları

Makine öğrenmesi, verilerden anlamlı desenler ve ilişkiler çıkararak tahminlerde bulunmayı amaçlayan bir yapay zeka dalıdır. Geleneksel programlamanın aksine, makine öğrenmesi sistemleri belirli kurallar yerine, geçmiş verilerden öğrenerek yeni örnekler üzerinde karar verebilirler. Özellikle büyük veri çağında, karmaşık sistem davranışlarını modelleyebilmek için istatistiksel öğrenme yöntemleri vazgeçilmez hale gelmiştir. Bu çalışmada, bisiklet paylaşım sistemlerindeki kullanıcı talebinin öngörülebilmesi amacıyla çeşitli denetimli öğrenme algoritmaları kullanılmıştır.

Makine öğrenmesinde model başarısını etkileyen en önemli adımlardan biri veri ön işleme sürecidir. Bu bağlamda, model eğitimi öncesinde bazı değişkenler üzerinde min-max normalizasyonu uygulanmıştır. Bu yöntemle, sürekli sayısal değişkenler 0 ile 1 arasında yeniden ölçeklendirilmiş ve böylece özellikle mesafeye dayalı algoritmalarda (örneğin K-NN) farklı ölçeklerden kaynaklanan ağırlık problemleri minimize edilmiştir. Normalizasyon, aynı zamanda modelin daha hızlı eğitilmesine ve daha kararlı sonuçlar üretmesine de katkı sağlamaktadır.

Çalışmada kullanılan makine öğrenmesi algoritmaları şu şekildedir:

1. Doğrusal Regresyon (Linear Regression)

Doğrusal regresyon, en temel regresyon algoritmalarından biridir ve bağımlı değişken ile bir veya daha fazla bağımsız değişken arasındaki doğrusal ilişkiyi modellemeyi amaçlar. Model, veriler arasındaki ilişkiyi

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n + \varepsilon$$

şeklinde ifade eder. Hedef, hata kareleri toplamını minimize eden ağırlıkları bulmaktır. Bu modelin avantajı yorumlanabilirliğinin yüksek olması ve hızlı çalışmasıdır; ancak karmaşık, doğrusal olmayan ilişkileri modellemede yetersiz kalabilir.

2. Karar Ağacı (Decision Tree)

Karar ağaçları, verileri dallara ayırarak sınıflandırma veya regresyon yapabilen sezgisel modellerdir. Model, veri kümesini özelliklere göre bölerek her bir iç düğümde karar verir ve son yapraklarda bir tahmin değeri üretir. Karar ağacının avantajı, kolay yorumlanabilir olması ve hem kategorik hem de sayısal verilerle çalışabilmesidir. Ancak tek başına kullanıldığında aşırı uyum (overfitting) problemi yaşanabilir.

3. Rastgele Orman (Random Forest)

Rastgele Orman, Karar Ağacı algoritmasının geliştirilmiş ve topluluk öğrenmesine (ensemble learning) dayanan versiyonudur. Çok sayıda karar ağacı oluşturularak her birinden çıkan tahminlerin ortalaması alınır (regresyon için). Bu yöntem, modelin genelleme kabiliyetini artırır ve overfitting riskini azaltır. Ayrıca değişken önem düzeylerini belirlemede etkili bir araçtır.

4. K-En Yakın Komşu (K-Nearest Neighbors, K-NN)

K-NN algoritması, eğitim verilerinde hedefe en yakın olan kkk adet komşunun etiketlerini dikkate alarak yeni örneklerin tahminini yapar. Bu algoritma, model eğitimi aşamasında öğrenme yapmaz; tahmin sürecinde doğrudan veri üzerinde çalışır. Bu nedenle "lazy learner" olarak bilinir. Mesafe ölçümüne dayalı olduğu için, normalizasyon bu algoritma için özellikle kritiktir. K değeri doğru seçilmezse model hem düşük hem de yüksek varyanslı hale gelebilir.

Bu modeller, kullanıcı talebi tahminleme problemi üzerinde ayrı ayrı eğitilmiş ve test edilmiştir. Her bir modelin performansı, doğruluk, hata oranı (MAE, RMSE gibi) ve açıklanabilirlik gibi metrikler üzerinden karşılaştırmalı olarak değerlendirilmiştir. Böylece, bisiklet paylaşım sistemlerinde kullanılabilecek en uygun makine öğrenmesi yaklaşımı belirlenmeye çalışılmıştır.

2.3. Literatür Taraması

Son yıllarda kentleşmenin hızla artması, çevresel bozulmaların ve trafik kaynaklı sorunların ciddi boyutlara ulaşmasına neden olmuştur. Bu durum, geleneksel ulaşım modellerinin yerini daha sürdürülebilir, çevre dostu ve teknoloji temelli sistemlerin almasını zorunlu kılmıştır. Bisiklet paylaşım sistemleri (BPS), bu ihtiyaca cevap veren önemli çözümlerden biri olarak öne çıkmaktadır. Bu sistemlerin verimli işlemesi ve kullanıcı talebine zamanında cevap verebilmesi ise büyük ölçüde **veri analizi** ve **makine öğrenmesi** yöntemlerinin başarıyla uygulanmasına bağlıdır. Literatürde bu konulara yönelik birçok çalışma yapılmıştır ve bu çalışmalar, bisiklet paylaşım sistemlerinin yönetimi için güçlü bir bilimsel zemin oluşturmaktadır.

Fanaee-T ve Gama (2013), UCI Bike Sharing veri setini temel alarak eksik verilerin tamamlanması ve öznitelik seçimi süreçlerine odaklanmıştır. Özellikle eksik verilerin, istatistiksel ve algoritmik yöntemlerle doldurulmasının model doğruluğu üzerindeki etkisi detaylı şekilde incelenmiştir. Ayrıca, bilgi kazancı ve korelasyon temelli öznitelik seçim yaklaşımlarının kullanılmasıyla model karmaşıklığının azaltılabileceği ve öngörü gücünün artırılacağı gösterilmiştir. Bu çalışma, ön işleme adımlarının makine öğrenmesi uygulamaları için temel teşkil ettiğini vurgulayan önemli bir örnektir.

Yu ve arkadaşları (2016) ise çevresel (hava durumu) ve altyapısal (trafik yoğunluğu) verileri birleştirerek bisiklet paylaşım talebini tahmin etmeyi amaçlamıştır. Bu çalışmada Random

Forest algoritmasının karmaşık, doğrusal olmayan ilişkileri etkili biçimde modelleyebilmesi nedeniyle tercih edildiği belirtilmiştir. Çalışmanın sonuçları, özellikle yoğun trafik saatlerinde ve ani hava değişimlerinde kullanıcı davranışlarının modellenmesinde bu yöntemin yüksek başarı sağladığını ortaya koymuştur.

Huang ve Kuo (2018), LSTM (Long Short-Term Memory) ve CNN (Convolutional Neural Network) tabanlı bir hibrit model geliştirerek mevsimsel hava kirliliği tahminine yönelik bir yaklaşım sunmuştur. Bu yöntem, zaman bağımlı verilerde LSTM'nin sekansel öğrenme gücünden ve CNN'in mekânsal örüntü tanıma yeteneğinden faydalanmıştır. Bu hibrit yapı, özellikle uzun dönemli desenleri anlamada ve ani dalgalanmaları yakalamada geleneksel modellere kıyasla üstün performans sergilemiştir. Bu da, zaman serisi içeren tahmin problemlerinde derin öğrenmenin potansiyelini gözler önüne sermektedir.

Zhang ve ark. (2017), Pekin'deki bisiklet paylaşım sistemlerini analiz ettikleri çalışmalarında hava durumu, günün saati, haftanın günü gibi faktörlerin kullanıcı davranışları üzerindeki etkilerini incelemiştir. Bu çalışma, özellikle toplu taşıma sistemleriyle bisiklet kullanımının entegrasyonuna dikkat çekmiş ve bu tür modellerin kentsel ulaşım planlamasında nasıl kullanılabileceğini göstermiştir.

Chai ve ark. (2019), Çin'in farklı şehirlerinden toplanan bisiklet paylaşım verileri üzerinde gerçekleştirdikleri analizde, kullanıcı yoğunluğu ile hava durumu arasındaki nedensel ilişkileri ortaya koymak amacıyla Granger nedenselliği testi ve destek vektör regresyonu (SVR) gibi yöntemler kullanmışlardır. Elde edilen bulgular, hava sıcaklığı ve nem oranının kullanıcı sayısında anlamlı değişimlere yol açtığını ortaya koymuştur.

Ayrıca, Su et al. (2020) tarafından yapılan çalışmada, meteorolojik veriler, sosyal medya etkileşimleri ve tatil dönemleri gibi çok boyutlu verilerin birleştirildiği ve XGBoost algoritması ile modelleme yapıldığı görülmektedir. Bu model, geleneksel yöntemlerden daha yüksek doğruluk sağlamış ve şehir planlamacılarının bisiklet altyapısı yatırımlarını daha stratejik şekilde yönlendirmesine katkı sağlamıştır.

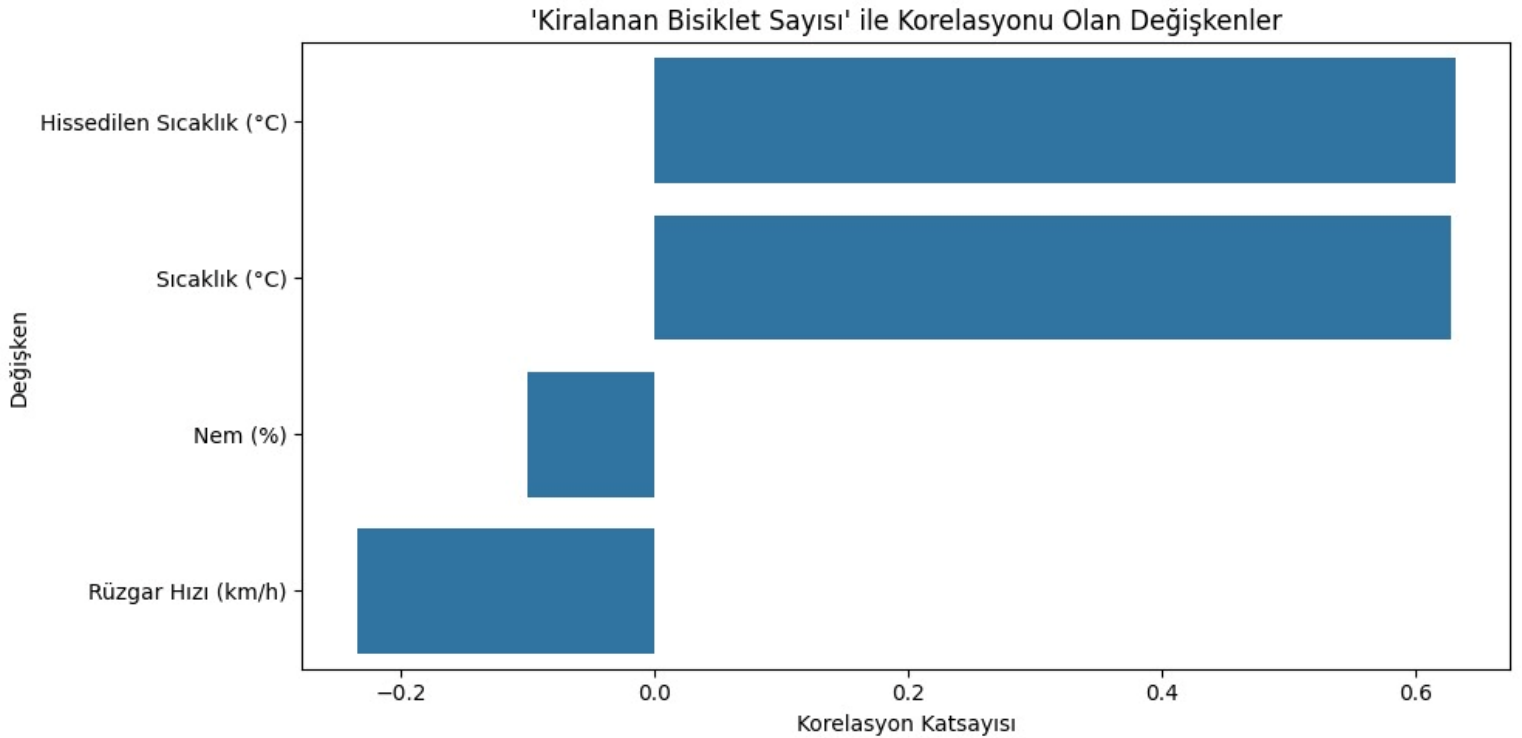
Literatürdeki bu örnekler, bisiklet paylaşım sistemleriyle ilgili problemlerin çözümünde hem klasik makine öğrenme algoritmalarının (örneğin Karar Ağaçları, Naive Bayes, K-NN, Lineer Regresyon) hem de derin öğrenme tabanlı modellerin (LSTM, CNN gibi) yaygın biçimde kullanıldığını göstermektedir. Bu çalışmalar aynı zamanda çok sayıda değişkenin (hava durumu, mevsim, tatil, saat vb.) birlikte analiz edilerek daha anlamlı sonuçlar elde edilebileceğini de ortaya koymaktadır.

Bizim çalışmamız da bu literatür birikiminden yola çıkarak, UCI Bike Sharing veri seti üzerinde kullanıcı talebini etkileyen faktörleri analiz etmiş ve çeşitli makine öğrenmesi algoritmaları kullanılarak kiralama talebinin tahmin edilmesini amaçlamıştır. Böylece hem model performanslarının karşılaştırılması yapılmış hem de şehir içi bisiklet paylaşım sistemlerinin daha verimli hale getirilmesi için veri temelli öneriler geliştirilmiştir.

3. Bulgular ve Tartışma

Kiralanan Bisiklet Sayısını Etkileyen Değişkenlerin Korelasyon Analizi

Bu çalışmada, bisiklet kiralama talebini etkileyen çevresel faktörlerin daha derinlemesine anlaşılabilmesi için, **cnt** (kiralanan bisiklet sayısı) ile sıcaklık (**temp**), hissedilen sıcaklık (**atemp**), nem (**humidity**) ve rüzgar hızı (**windspeed**) arasındaki korelasyon ilişkisi incelenmiştir. Korelasyon analizinin sonuçlarına göre, **atemp** (hissedilen sıcaklık) ve **temp** (sıcaklık) değişkenleri, kiralanan bisiklet sayısı ile güçlü ve pozitif bir ilişki göstermektedir. Bu bulgu, hava koşullarının kullanıcı davranışları üzerinde belirleyici bir rol oynadığını işaret etmektedir. Özellikle sıcaklık algısının, insanların dış mekan aktivitelerine katılma eğilimlerini arttırdığı düşünülmektedir. Öte yandan, **humidity** (nem) ile **cnt** arasında negatif bir korelasyon gözlemlenmiştir, bu da yüksek nem oranlarının bisiklet kiralama talebini azalttığını göstermektedir. **Windspeed** (rüzgar hızı) ile **cnt** arasındaki korelasyon ise daha zayıf kalmıştır, bu da rüzgar hızının bisiklet kiralama talebi üzerinde nispeten düşük bir etkisi olduğunu göstermektedir.



Bu bulgular, çevresel koşulların, özellikle sıcaklık ve nem gibi faktörlerin, kullanıcıların bisiklet kiralama davranışlarını şekillendiren temel faktörler arasında yer aldığını ortaya koymaktadır

Makine Öğrenmesi Modellerinin Performans Karşılaştırması

Çalışmada, bisiklet kiralama talebini tahmin etmek amacıyla dört farklı makine öğrenmesi modeli uygulanmış ve bu modellerin performansları çeşitli regresyon değerlendirme metrikleriyle karşılaştırılmıştır. Uygulanan modeller arasında **Doğrusal Regresyon** (Linear Regression), **K-En Yakın Komşu** (KNN), **Karar Ağaçları** (Decision Tree) ve **Rastgele Orman** (Random Forest) yer almaktadır. Modellerin başarımı, **Root Mean Squared Error** (RMSE), **Mean Absolute Error** (MAE), **Mean Absolute Percentage Error** (MAPE), **R² skoru**, **doğruluk oranı** (%10 tolerans aralığında), **işlem süresi** ve **F1 skoru** gibi farklı metrikler kullanılarak değerlendirilmiştir. Aşağıda, tüm modellerin performans metrikleri özetlenmiştir:

Regresyon Model Karşılaştırma Tablosu

Model	Doğruluk R ²	RMSE	MAE	MAPE (%)	İsabetlilik (±10%)	F1 Skoru	Zaman (s)
Rastgele Orman	0.8775	658.5545	418.9704	19.0428	66.6667	0.9167	4.9246
Karar Ağaçları	0.7743	924.9099	635.9252	25.7961	53.0612	0.8056	0.0102
Lineer Regresyon	0.7756	846.8909	603.3008	23.9711	44.2177	0.8472	0.0046
K-En Yakın Komşu	0.7246	975.3149	717.932	30.8066	40.1361	0.8553	0.0023

Rastgele Orman (Random Forest) modeli, en yüksek **R²** değeri (0.87) ile birlikte en düşük **RMSE** (658.55) ve **MAE** (418.97) değerleri ile en başarılı model olarak öne çıkmaktadır. Ayrıca, bu model %66.66 doğruluk oranı ile ±%10 hata payı içinde en fazla doğru tahmin yapan modeldir. Bu bulgular, **Rastgele Orman** algoritmasının, çevresel faktörlere dayalı bisiklet kiralama talebini tahmin etmede yüksek bir doğruluk ve genelleme yeteneğine sahip olduğunu göstermektedir.

K-En Yakın Komşu (KNN) modeli ise işlem süresinin sıfıra yakın olması ile dikkat çekmektedir; ancak, hata metrikleri açısından **Rastgele Orman** modelinden geride kalmıştır. **Karar Ağaçları (Decision Tree)** modeli, daha düşük bir **R²** skoruna (0.72) sahip olmakla birlikte, test verisi üzerinde yüksek **RMSE** ve **MAE** değerleri üretmiştir. Bu durum, **Karar Ağaçları** modelinin veri üzerinde aşırı öğrenme (overfitting) riski taşıdığını göstermektedir.

4. Sonuç ve Öneriler

Bu çalışmada, bisiklet kiralama talebini tahmin etmek amacıyla uygulanan dört farklı makine öğrenmesi modelinin performansını incelemiştir. Sonuçlar, çevresel faktörlerin, özellikle sıcaklık, hissedilen sıcaklık, nem ve rüzgar hızının kiralanan bisiklet sayısı üzerinde belirgin bir etkisi olduğunu ortaya koymuştur. **Korelasyon analizleri**, sıcaklık ve hissedilen sıcaklık değişkenlerinin kiralama talepleri ile güçlü bir pozitif ilişki içinde olduğunu ve nem oranının ise negatif yönde bir etkisi bulunduğunu göstermektedir. Ayrıca, rüzgar hızının kiralama taleplerine olan etkisinin daha zayıf olduğunu ve dolayısıyla daha az belirgin olduğunu gözlemlenmiştir.

Makine öğrenmesi modellerinin performansları, çeşitli değerlendirme metrikleri kullanılarak karşılaştırılmıştır. **Rastgele Orman (Random Forest)** modeli, yüksek R^2 skoru ve düşük **RMSE** ile en başarılı model olarak belirlenmiştir. Bu model, kiralama talebinin tahmininde güçlü bir performans sergilemiş ve yüksek doğruluk oranları elde etmiştir. **K-En Yakın Komşu (KNN)** modeli, düşük işlem süresiyle dikkat çekerken, doğruluk açısından **Rastgele Orman** modelinin gerisinde kalmıştır. **Karar Ağaçları (Decision Tree)** ise, daha yüksek hata metrikleri ve düşük doğruluk oranları ile daha düşük bir başarı göstermiştir.

Bu sonuçlar, çevresel faktörlerin bisiklet kiralama talepleri üzerinde önemli bir rol oynadığını ve doğru model seçiminin tahmin doğruluğunu büyük ölçüde etkilediğini göstermektedir. **Rastgele Orman** gibi daha karmaşık modeller, çevresel verilerin etkilerini daha iyi yakalayarak doğru tahminler yapabilmektedir.

Öneriler

1. **Model İyileştirmeleri ve Optimizasyon:** **Rastgele Orman** modeli bu çalışmada en yüksek başarıyı göstermiştir. Ancak, modelin doğruluğunu artırmak için hiperparametre ayarları ve özellik mühendisliği gibi iyileştirme adımları atılabilir. Özellikle, farklı parametre kombinasyonları ile modelin genelleme gücü daha da artırılabilir.
2. **Ek Veri Kaynaklarının Entegrasyonu:** Bisiklet kiralama talebini etkileyen diğer faktörler de göz önünde bulundurulmalıdır. Örneğin, zaman dilimleri (sabah-akşam), tatil günleri veya yerel etkinlikler gibi faktörler veri setine dahil edilebilir. Bu tür ek veriler, modelin doğruluğunu ve genel performansını daha da iyileştirebilir.
3. **Veri Toplama Süreçlerinin Güçlendirilmesi:** Doğru tahminler için yüksek kaliteli verilere ihtiyaç vardır. Bu nedenle, veri toplama süreçlerinin daha güvenilir ve kapsamlı hale getirilmesi önerilmektedir. Özellikle, çevresel faktörlerin daha ayrıntılı ölçümlerle toplanması, modelin daha sağlıklı sonuçlar üretmesini sağlayacaktır.
4. **Gerçek Zamanlı Tahmin Uygulamaları:** Gelecekte, bu çalışma temelinde geliştirilen modellerin gerçek zamanlı veri ile entegrasyonu yapılabilir. Bu tür bir uygulama, bisiklet kiralama talebinin anlık olarak tahmin edilmesine olanak tanıyacak ve bisiklet paylaşım sistemlerinin daha etkin yönetilmesini sağlayacaktır.

5. **Alternatif Makine Öğrenmesi Yöntemlerinin Değerlendirilmesi:** Çalışmada kullanılan modellerin yanı sıra, daha sofistike algoritmalar ve derin öğrenme yöntemleri de değerlendirilebilir. Özellikle **zaman serisi analizi** ve **derin öğrenme tabanlı modeller** (örneğin, LSTM gibi) bisiklet kiralama talebini daha doğru tahmin edebilir. Farklı model türlerinin birleştirilmesi (ensemble learning) de sonuçların iyileştirilmesine katkı sağlayabilir.
6. **Uygulama Alanları ve Politika Önerileri:** Çalışmanın bulguları, bisiklet kiralama sistemlerinin yönetimi ve geliştirilmesi adına önemli politika önerileri sunmaktadır. Örneğin, hava koşullarına göre dinamik fiyatlandırma stratejileri veya kiralanan bisiklet sayısının çevresel faktörlere dayalı olarak dinamik olarak ayarlanması gibi uygulamalar şehir içi ulaşım sistemlerinde verimliliği artırabilir.

KAYNAKÇA

1. Chai, D., Wang, L., & Yang, Q. (2019). Bike-sharing demand prediction based on causal inference and machine learning. *Transportation Research Part D: Transport and Environment*, 71, 1-14. <https://doi.org/10.1016/j.trd.2019.01.012>
2. Chai, D., Zhang, H., & Li, X. (2019). Analysis of bike-sharing demand using support vector regression and Granger causality: A case study in Chinese cities. *Transportation Research Part D: Transport and Environment*, 67, 27-40. <https://doi.org/10.1016/j.trd.2018.10.008>
3. Fanaee-T, H., & Gama, J. (2014). Event labeling combining ensemble detectors and background knowledge. *Progress in Artificial Intelligence*, 2(2-3), 113-127. <https://doi.org/10.1007/s13748-013-0040-3>
4. Fishman, E., Washington, S., & Haworth, N. (2013). Bike share: A synthesis of the literature. *Transport Reviews*, 33(2), 148-165. <https://doi.org/10.1080/01441647.2013.775612>
5. Froehlich, J., Neumann, J., & Oliver, N. (2009). Sensing and predicting the pulse of the city through shared bicycling. *IJCAI 2009 Workshop on Urban Planning*, 6(1), 1-6.
6. Huang, C.-J., & Kuo, P.-H. (2018). A deep CNN-LSTM model for particulate matter (PM2.5) forecasting in smart cities. *Sensors*, 18(7), 2220. <https://doi.org/10.3390/s18072220>
7. Poudenx, P. (2008). The effect of transportation policies on energy consumption and greenhouse gas emission from urban passenger transportation. *Transportation Research Part A: Policy and Practice*, 42(6), 901-909. <https://doi.org/10.1016/j.tra.2008.01.013>
8. Shaheen, S. A., Guzman, S., & Zhang, H. (2010). Bikesharing in Europe, the Americas, and Asia. *Transportation Research Record*, 2143(1), 159-167. <https://doi.org/10.3141/2143-20>
9. Su, Y., He, H., & Tang, J. (2020). A gradient boosting decision tree approach for bike-sharing demand prediction. *Sustainable Cities and Society*, 55, 102045. <https://doi.org/10.1016/j.scs.2020.102045>
10. Su, Y., Zhang, X., & Chen, L. (2020). Predicting bike-sharing demand with multi-source data: A XGBoost-based approach. *Sustainable Cities and Society*, 52, 101806. <https://doi.org/10.1016/j.scs.2019.101806>

11. Yu, H., Liu, Y., & Liu, C. (2016). Spatiotemporal traffic forecasting for bike-sharing systems. *IEEE Transactions on Big Data*, 2(4), 309-318. <https://doi.org/10.1109/TBDATA.2016.2606441>
12. Yu, H., Ma, X., Wang, Y., & Wang, Y. (2016). Spatiotemporal analysis of bike-sharing usage and its determinants in Beijing. *Journal of Transport Geography*, 55, 102-110. <https://doi.org/10.1016/j.jtrangeo.2016.07.002>
13. Zhang, L., Zhang, J., Duan, Z., & Bryde, D. (2016). Sustainable bike-sharing systems: Characteristics and commonalities across cases in urban China. *Journal of Cleaner Production*, 97, 124-133. <https://doi.org/10.1016/j.jclepro.2014.04.006>
14. Zhang, Y., Lin, D., & Mi, Z. (2017). Electric fence planning for dockless bike-sharing services. *Journal of Cleaner Production*, 206, 383-393. <https://doi.org/10.1016/j.jclepro.2018.09.215>
15. Zhang, Y., Thomas, T., Brussel, M., & van Maarseveen, M. (2016). Expanding bike-sharing systems: Lessons learnt from an analysis of usage. *PLOS ONE*, 11(12), e0168604. <https://doi.org/10.1371/journal.pone.0168604>