

Havayolu Taşımacılığında Müşteri Memnuniyetinin Topluluk Öğrenmesi Yöntemleri ile Belirlenmesi

(Araştırma Makalesi)

Determining Airline Customer Satisfaction with Ensemble Learning Methods

Doi: 10.29023/alanyaakademik.1095574

Uğur ERCAN

Öğr. Gör. Dr., Akdeniz Üniversitesi, Enformatik Bölüm Başkanlığı, Antalya/Türkiye

ugurercan@akdeniz.edu.tr,

Orcid No: 0000-0002-9977-2718

Bu makaleye atıfta bulunmak için: Ercan, U. (2022). Havayolu Taşımacılığında Müşteri Memnuniyetinin Topluluk Öğrenmesi Yöntemleri ile Belirlenmesi. Alanya Akademik Bakış, 6(3), Sayfa No.2763-2774.

ÖZET

Anahtar kelimeler:
Havayolu müşteri memnuniyeti, Rastgele Orman, Gradient Boosting, XGBoost, havayolu taşımacılığı.

Makale Geliş Tarihi:
30.03.2022
Kabul Tarihi:
15.09.2022

Havayolu taşımacılığında seyahatinden memnun olmayan müşterilerin tahmin edilmesi firmaların kendilerini yapılandırması ve gelirlerinin yönetilmesi açısından son derece önemlidir. Gerçekleştirilen çalışmada Amerika Birleşik Devletleri'ndeki havayollarından derlenen veriler kullanılarak uçuş seyahatinden nötr ya da memnun olmayan müşterilerin topluluk öğrenmesi yöntemleriyle tahmin edilmesi amaçlanmıştır. Modelleme aşamasında sınıflandırma problemlerinde yüksek tahmin doğruluğu üreten ve güncel makine öğrenmesi yöntemlerinden Rastgele Orman, Gradient Boosting ve XGBoost yöntemleri kullanılmıştır. Elde edilen en iyi doğruluk oranı %96,4 iken en iyi Özgüllük ve Negatif Tahmin Oranı değerleri sırasıyla %97,7 ve %96'dır. Model sonuçlarından elde edilen yüksek Özgüllük, Negatif Tahmin Oranı ve Doğruluk değerleri makine öğrenmesi yöntemlerinin havayolu taşımacılığında müşterilerin havayolu firmasını tekrar kullanıp kullanmayacağı tahmin işlemlerinde kullanılabileceğini göstermektedir.

ABSTRACT

Keywords:
Airline customer satisfaction, random forest, Gradient Boosting, XGBoost, airline transportation.

Estimating the customers who aren't satisfied with their travels in air transportation is extremely important in terms of structuring the companies themselves and managing the revenues. In this study, it's aimed to predict neutral or dissatisfied customers with ensemble learning methods by using data compiled from airlines in the U.S. In the modeling phase, Random Forest, Gradient Boosting, and XGBoosting methods, which are current machine learning methods that produce high estimation accuracy in classification problems, were used. The best accuracy obtained was 96.4%, while the best Specificity and Negative Prediction Rate values were 97.7% and 96%, respectively. The high Specificity, Negative Prediction Rate, and Accuracy values obtained from the model results show that machine learning methods can be used to predict whether customers will reuse airline companies in airline transportation.

1. GİRİŞ

Ulaştırma sektörünün önemli bir alt sektörü olan havayolu ulaştırması; faaliyet konusu, faaliyeti yürüten kamu ve özel kurum ve kuruluşlar, kullanılan yüksek teknoloji ürünü araçlar ve donanım, özel altyapı ve haberleşme sistemleri, nitelikli insan gücü, hizmet verilen insanlar, ulusal ve uluslararası özelliğe sahip kurallar ve mevzuat konularının oluşturduğu önemli bir sistemdir (DPT, 2001: 1). Bir dönem pahalı olması sebebi ile lüks olarak görülen, az sayıda insan tarafından kullanılan uçak yolculuğu son yıllarda bilet fiyatlarının düşmesi ile ciddi bir artış göstermiştir (Küçük Çırpın ve Kurt, 2016: 84). Havayolu endüstrisi, en rekabetçi hizmet sektörlerinden biri olarak kabul edildiğinden, birçok firma, müşterilerinin firmalarına olan sadakatini artırmayı ve böylece müşterilerin firmalarını tekrar tercih etmesini amaçlamaktadır (Hwang vd., 2020: 122). Havayolu firmalarının sayısının artması ile birlikte rekabet ön plana çıkmakta, böylece firmaların sağladığı hizmetlerin kalitesi ve fiyatları firmaların tercih edilmesinde majör bir rol oynamaktadır. Firmaların havayolu pazarında varlıklarını devam ettirmeleri, pazar payını artırmaları ancak rekabetçi bir anlayışla müşteri memnuniyetini istenilen düzeye getirebilmeleriyle mümkün olabilecektir (Çelikkol vd., 2012: 71). Uçuş sırasında müşterilerin ihtiyaçlarını ve konfor düzeyini, yani müşteri memnuniyetini anlamak çok önemlidir. Bu nedenle, müşteri geri bildirimini herhangi bir havayolu endüstrisi için çok önemlidir (Kumar ve Zymbler, 2019: 2). Mal ve hizmet satın alırken hem kuruluşlar hem de onları tedarik eden kişiler için önceki müşterilerin deneyim, görüş ve önerileri çok önemlidir (Baydoğan ve Alatas, 2019: 1). Benzer şekilde havayolu endüstrileri içinde müşteri deneyimi önemli bir endişe kaynağıdır (Kumar ve Zymbler, 2019: 1). Çünkü firmalar hakkında yapılan kötü yorumların yaygınlaşması bu firmalar için büyük bir felaketin habercisidir (Baydoğan ve Alatas, 2019: 1). Yolcular herhangi bir havayolu firmasını seçmeden önce bilet ücreti, seyahat süresi, duraklama sayısı, izin verilen bagaj sayısı ve mevcut müşteri geri bildirim gibi birkaç noktayı göz önünde bulundurur (Kumar ve Zymbler, 2019: 2). Tüm bunların yanında hizmet kalitesi değişkenleri de müşteri memnuniyet düzeylerini etkileyen önemli faktörlerdendir (Jiang ve Zhang, 2016: 80). Bu nedenle havayolu firmaları, tesislerini ve uçuş konforlarını iyileştirmek böylece müşterileri çekmek için müşteri hizmetleri alanında yoğunlukla çalışmaktadır (Kumar ve Zymbler, 2019: 2).

Gerçekleştirilen bu araştırmanın amacı, havayolu uçuşundan nötr ya da memnun olmayan müşterilerin güncel topluluk tabanlı makine öğrenmesi yöntemlerinden olan Random Forest, XGBoost, Gradient Boosting ile tahmin edilmesi, elde edilen model sonuçlarına göre model başarımlarının karşılaştırılmasıdır.

2. KAVRAMSAL ÇERÇEVE

Literatürde gerçekleştirilen çalışmalar incelendiğinde, An ve Noh (2009) uçak içi hizmet kalitesinin havayolu müşteri memnuniyeti ve sadakati üzerindeki etkisini araştırmıştır. Prestij sınıfı için alkollü ve alkolsüz içecekler, hızlı çözüm oluşturma ve empati, güvenilirlik, sigorta, yemeğin sunum tarzı ve kalitesi; ekonomi sınıfı için hızlı çözüm oluşturma ve empati, yiyecek kalitesi, alkollü ve alkolsüz içecek, güvenilirliğin müşteri memnuniyeti ve sadakatini etkilediğini belirtmiştir. Çelikkol vd. (2012) havayolu taşımacılığında müşteri tercihini ve memnuniyetini etkileyen faktörleri belirlemiştir. Müşterilerin tercihlerine ve memnun olmalarını sağlayan faktörlerin güvenilirlik ve iyi hizmet, uzmanlık ve esneklik, avantaj ve uygunluk, güvenlik ve kalite, rahatlık ve çeşitlilik olduğu görülmüştür. Hussain vd. (2015) havayolunda hizmet kalitesi, hizmet sağlayıcı imajı, müşteri beklentileri, algılanan değer, müşteri memnuniyeti ve marka sadakati arasındaki bağlantıları araştırmıştır. Hizmet kalitesi, 2764

algılanan değer ve marka imajının müşteri memnuniyeti üzerinde olumlu ve anlamlı bir etkiye sahip olduğunu ve bunun da marka sadakatine yol açabileceğini belirtmiştir. Jiang ve Zhang (2016) bilet fiyatlandırması, cinsiyet, gelir, eğitimin istatistiksel olarak anlamlı ve yolcuların genel memnuniyeti üzerinde olumlu ve önemli bir etkisi olduğunu belirtmiştir. Giao (2017) yolcu memnuniyetine etki eden faktörleri belirlemiştir. Müşteri memnuniyetine etki eden faktörlerin; uçağa biniş/iniş/bagaj, check-in, uçak içi hizmetler, rezervasyon, uçak ve uçuş ekibi olduğunu belirtmiştir. Cinsiyet, yaş grupları, aylık gelir grupları ve eğitim grupları arasında memnuniyet farkı bulunmazken, yolcuların seyahat amaçları (iş, konferans, turizm, vd.) arasında fark bulunduğu görülmüştür. Baydoğan ve Alatas (2019) doğal dil işleme ve makine öğrenmesi yöntemleri ile havayolu firmalarına ait ürün veya hizmetlere yapılan yorumlardan müşterinin memnuniyet durumunu tahmin etmiştir. Sıralı Minimum Optimizasyon (Sequential Minimal Optimization) ve Çok Sınıflı Sınıflandırma (Multi Class Classification) yöntemlerinin doğruluk değerleri uygulandıkları çoğu veri setinde diğer yöntemlere göre daha başarılı olduğu görülmüştür. Topal vd. (2019) havayolu yolcularının, şirketlerin vermiş olduğu hizmetlerden algıladıkları hizmet kalitesinin müşteri memnuniyeti ve müşterilerin yeniden satın alma davranışına yansımalarını belirlemiştir. Cinsiyet, uçuş sınıfı, uçuş yönü, gelir ve meslek grubunun memnuniyeti etkilediği belirtilmiştir. Hwang vd. (2020) daha önce aldıkları hizmet kullanımına ait alınan geri bildirim yorumları ve memnuniyet derecelendirmeleri üzerinden makine öğrenimi yaklaşımı ile müşterilerin havayolu hizmetlerini tekrar ziyaret kullanma olasılığını tahmin etmiştir. Müşteriler tarafından yapılan yorumlardaki duygusal özellikler göz önüne alındığında, sınıflandırma yöntemlerinin, müşterilerin tekrar ziyaretlerini tahmin etmede %83,42 doğruluk gösterdiği görülmüştür. Ayrıca, müşteriler tarafından yazılan geri bildirimlerin daha yüksek kelime sayısına sahip olmasının tahmin doğruluğunu artırabileceği belirtilmiştir. Vinod (2021) müşteri memnuniyeti için yapay zekâ ve makine öğrenmesi yöntemlerinin kullanılabilirliğini incelemiştir. Agarwal ve Govda (2021) uçak içi ortam koşulları, mürettebatın ilişki kalitesi ve diğer hizmetler açısından yolcuların memnuniyet düzeyini araştırmıştır. Yolcuların, uçak içi hizmet, uzaktan uçak içi hizmet ve arka ofis operasyonlarında sunulan hizmetlerin kalitesinden memnun olduğu görülürken, eğitilmiş ve yüksek gelirli yolcuların daha sık kullanılan uçuş kullandığı görülmüştür. Yolculara kaliteli hizmetin sunulmamasının havayolunun imajına zarar verebileceği ve yolcuların davranışsal niyetleri üzerinde olumsuz etkilere neden olabileceği belirtilmiştir. Gao vd. (2021) havayolu seyahat memnuniyetini belirlemek için makine öğrenmesi yöntemlerini kullanmışlardır. Rastgele Orman yöntemi ile kurulan model tüm ölçütlerde en iyi sonuçları elde etmiştir. Noviantoro ve Huang (2022) makine öğrenmesi yöntemleri ile havayolu yolcularının memnuniyetini incelemiştir. Derin öğrenme, Karar Ağaçları, Lojistik Regresyon, Yapay Sinir Ağları, Kural İndüksiyonu (Rule Induction) Destek Vektör Makineleri, k-NN, Naive Bayes ve Gradient Boosting modellerinin kullanıldığı çalışmada Derin Öğrenme ile kurulan model bütün metriklerde en iyi sonuçları elde etmiştir. Yolcu memnuniyetini kazanmak için çevrimiçi/mobil boarding, uçak içi wi-fi hizmeti, bagaj taşıma ve uçak içi eğlencenin havayolu firmaları tarafından geliştirilmesi gereken en önemli dört hizmet olduğu belirtilmiştir.

3. VERİ SETİ

Araştırmada veri seti olarak açık erişimli, ücretsiz ve ikincil verilerin olduğu “www.kaggle.com” web sitesinde yer alan ve U.S. Airlines 2015 yılı uçuşlarından derlenen “Airline Passenger Satisfaction” verisi kullanılmıştır (Kaggle, 2022). Modelleme aşamasında kullanılan değişkenler ve tiplerine ait açıklamalar Tablo 1’de gösterilmiştir.

Tablo 1. Değişken Tanımları

Değişken Adı	Tipi	Açıklama
Gender	Kategorik	Müşterinin cinsiyeti (kadın, erkek)
CustomerType	Kategorik	Müşteri tipi (sadık ve sadık olmayan müşteri)
Age	Sayısal	Müşterinin yaşı
TypeofTravel	Kategorik	Uçuş tipi (İş seyahati, kişisel seyahat)
Class	Kategorik	Uçuş sınıfı (business, ekonomik, ekonomik plus)
FlightDistance	Sayısal	Uçuş mesafesi
InflightWifiService	Ordinal	Uçak içi wifi hizmet kalitesi
Departure/ArrivalTimeConvenient	Ordinal	Uçağın kalkış/varış saat uygunluğu
EaseOfOnlineBooking	Ordinal	Online rezervasyon kolaylığı
GateLocation	Ordinal	Kapı yeri
FoodAndDrink	Ordinal	Yiyecek ve içecek
OnlineBoarding	Ordinal	Online biniş
SeatComfort	Ordinal	Koltuk rahatlığı
InflightEntertainment	Ordinal	Uçak içi eğlence
On-BoardService	Ordinal	On Board servisi
BaggageHandling	Ordinal	Bagaj taşıma servisi
CheckinService	Ordinal	Checkin servisi
InflightService	Ordinal	Uçak içi servis
Cleanliness	Ordinal	Temizlik
DepartureDelayInMinutes	Sayısal	Kalkış gecikmesi (dakika)
ArrivalDelayInMinutes	Sayısal	Varış gecikmesi (dakika)
Satisfaction* (Hedef Değişken)	Kategorik	Uçuş yolculuğundan memnun olmuşsa (1), nötr ya da memnun olmamışsa (0)

*Ordinal veriler: 0-5 arasında puanlanmıştır. 0: Hizmetin bulunmamasını, 1: en düşük değeri, 5: en yüksek değeri göstermektedir.

Kaynak: Kaggle, 2022

Veri setinde toplam 129.880 gözlem yer almakta, her bir gözlem bir uçuş kaydını göstermektedir. 56.428 yolcu gerçekleştirdiği uçuştan memnun iken 73.452 kişi uçuş seyahatinden nötr ya da memnun değildir. Veri setinde, 21 bağımsız ve 1 çıkış değişkeni olmak üzere toplam 22 değişken yer almaktadır.

4. YÖNTEM

Makine öğrenmesi, örnek verileri veya geçmiş deneyimleri kullanarak sürekli bir çıktının değerinin ya da kategorik bir çıktının sınıfının tahmin edilmesi süreci olarak tanımlanabilir (Alpaydın, 2010). Makine öğrenmesi yöntemleri sağlık, eğitim, mühendislik, sosyal ve ziraat gibi birçok alandaki regresyon ve sınıflandırma problemlerinde sıklıkla kullanılmaktadır (Ecer vd., 2018; Mariescu-Istodor ve Jormanainen, 2019; Kumar vd., 2020; Rajula vd., 2020; Duman vd., 2022; Ercan, 2021). Gerçekleştirilen çalışmada güncel makine öğrenmesi yöntemleri sivil

hava ulaştırma alanına uygulanmıştır. Amerika Birleşik Devletleri'ndeki havayollarını kullanan müşterilerin seyahatlerinden memnun olma durumlarının tahmin edilmesine yönelik bir uygulama gerçekleştirilmiştir. Rastgele Orman, Gradient Boosting, XGBoost ve LightGBM makine öğrenmesi yöntemleri, Topluluk Öğrenme yöntemleridir. Bu yöntemler özellikle sınıflandırma problemlerinde etkin ve başarılı bir şekilde kullanılmaktadır (Ercan, 2021; Üstüner vd., 2021; Noviantoro ve Huang, 2022). Popüler topluluk öğrenme yöntemlerinden Gradient Boosting (GB), XGBoost ve Rastgele Orman (RO) ile modellemeler yapılmış ve sonuçları karşılaştırmalı olarak incelenmiştir. Çalışmada kullanılan veri seti, dengeli bir seti olduğu için SMOTE, ROSE gibi veri dengeleme yöntemleri kullanılmasına gerek kalmamıştır. Buna karşın modellerin doğruluğunu arttırmak için k-katlı çapraz doğrulama yöntemi kullanılmıştır. Model başarımlarının karşılaştırılmasında Doğruluk, Kesinlik, Duyarlılık, Özgüllük ve Negatif Tahmin Oranı metrikleri kullanılmıştır.

Gradient Boosting

Sınıflandırma ve regresyon problemlerinde etkin bir şekilde kullanılan Gradient Boosting (GB) (Flores ve Keith, 2019: 4) Fricadman (2001) tarafından geliştirilmiştir. GB'de amaç daha güçlü bir tahminci elde etmek için zayıf tahmincileri iterasyonlarla kademeli olarak daha güçlü tahmincilere dönüştürmektir (Ponraj ve Vigneswaran, 2020: 5738). GB algoritmasının amacı matematiksel olarak aşağıdaki şekilde ifade edilebilir. Bir eğitim veri kümesi $D\{x_i, y_i\}_1^N$ verildiğinde, belirli bir kayıp fonksiyonu $L(y, F(x))$ 'in beklenen değerini en aza indirerek, x örneklerini y çıktı değerlerine eşleyen, $F^*(x)$ fonksiyonunun bir yaklaşıklığı $\hat{F}(x)$ 'i bulmaktır. GB, $F_m(x) = F_{m-1}(x) + p_m h_m(x)$ fonksiyonlarının ağırlıklı toplamı olarak, $F^*(x)$ 'in toplamsal bir yaklaşıklığını oluşturur (Bentéjac vd., 2021: 1940). Bu yöntemde model, diğer Boosting yöntemlerinde olduğu gibi aşamalar halinde oluşturur, ancak bunları keyfi türevlenebilir bir kayıp fonksiyonunu optimize ederek geliştirir (Flores ve Keith, 2019: 4). Dolayısıyla GB algoritmasının çalışma mantığı bir sonraki tahmincinin önceki tahmin edicilerden öğreneceği mantığı kullanmasıdır (Anggraina vd., 2019: 1-2). Topluluk öğrenme metodları tuning adı verilen parametre ayarlamalarına karşı Yapay Sinir Ağlarına göre daha az duyarsız iken (Bustillo vd., 2011: 521) genellikle yorumlanmaları güçtür (Van Assche ve Blockeel, 2007: 418).

Extreme Gradient Boosting

Chen ve Guestrin (2016) tarafından geliştirilen, GB'a dayalı bir makine öğrenimi yöntemi olan eXtreme Gradient Boosting (XGBoost) yöntemi, optimal bir model oluşturmak için karar ağacında gradyan inişi kullanır (Rumora vd., 2020: 7). Topluluk öğrenme tabanlı bir yöntem olan XGBoost, yüksek performanslı karar ağacı modellerinin etkisini göz önünde bulundurarak önceki karar ağaçları tarafından yapılan hataların ardışık karar ağaçları ile en aza indirilmeye çalışılacağı şekilde sırayla birden fazla karar ağacı oluşturur. Bu işlem için Gradyan İniş (Gradient Descent) yöntemini kullanır (Kumar vd., 2022: 3). GB yönteminin sınırlamalarından kurtulmak için geliştirilen (Al Daoud, 2019: 1) XGBoost yönteminin yeniliği, model karmaşıklığını kontrol eden bir amaç fonksiyonu içermesidir. Bu amaç fonksiyonu, kayıp fonksiyonu ve düzenleme tekniğinin birleşiminden oluşmaktadır. Bu da paralel hesaplamaları ve optimum hesaplama hızının korunmasını mümkün kılar (Abdi, 2020: 7). XGBoost'da ağaçların karmaşıklığını kontrol etmek için kayıp fonksiyonunun bir varyasyonu kullanılır.

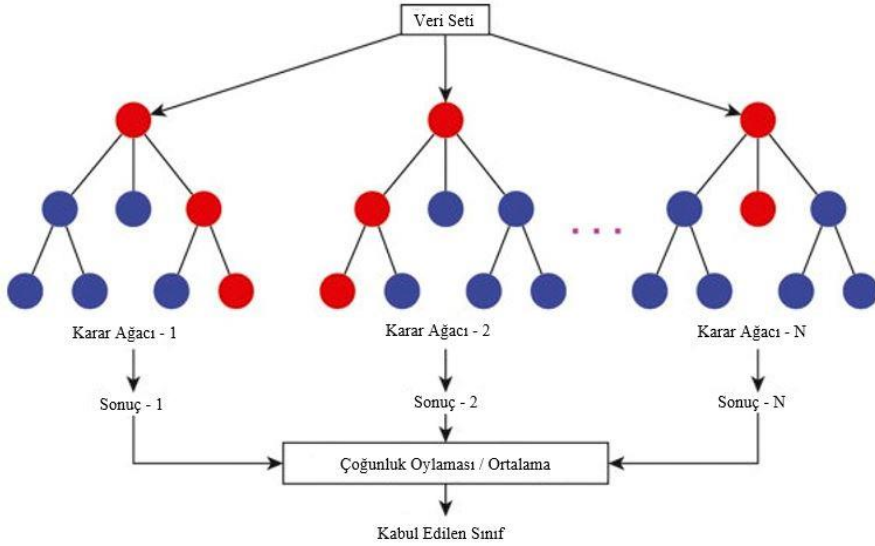
$$L_{xgb} = \sum_{i=1}^N L(y_i, F(x_i)) + \sum_{m=1}^M \Omega(h_m)$$

$$\Omega(h) = \gamma T + \frac{1}{2} \lambda \|w\|^2$$

T , ağacın yaprak sayısı ve w ise yaprakların çıktığı puanlarıdır. γ 'nin daha yüksek değerlere sahip olması daha sade ağaçlar elde edilmesine yol açar (Bentéjac vd., 2021: 1941-1942). XGBoost yönteminin başarısının temel nedeni, öğrenme sürecinde kullandığı amaç fonksiyonudur. Kayıp fonksiyonunun görevi, modelin yaptığı her bir tahmin değerinin gerçek değerinden olan farkını hesaplamaktır (Üstüner vd., 2020: 100). Düzenleştirme tekniği aşırı uyumu kontrol için kullanılır bu nedenle XGBoost, model ayarlama (tuning) sırasında hesaplama açısından çok hızlı ve daha sağlamdır (Rumora vd., 2020: 7). Düzenleştirme tekniği kayıp fonksiyonuna yeni bir terim eklenerek yapılır. Önceden oluşturulmuş model kullanılarak düzeltilen çoklu modeller üretir ve nihai modeli oluşturur (Al Daoud, 2019: 1).

Rastgele Orman

En yaygın bilinen topluluk öğrenme yöntemleri; Bagging, Boosting ve Rastgele Orman'dır (Oshiro vd., 2012: 154). Breiman (2001) tarafından geliştirilen topluluk tabanlı bir sınıflandırıcı olan Rastgele Orman (RO) yöntemi sınıflandırma ve regresyon problemlerinde sıklıkla kullanılmaktadır. Bagging yöntemine ait entegre bir algoritma olan RO yönteminde amaç, ağırlıklı veya ağırlıksız oylama yoluyla bir dizi sınıflandırıcının kararını birleştirmektir (Pal, 2005: 217; Zhang vd., 2020: 9). RO yönteminin çalışma prensibi Şekil 1'de gösterilmiştir.



Şekil 1. Rastgele Orman Yönteminin Çalışma Prensibi (Khan vd., 2021: 7)

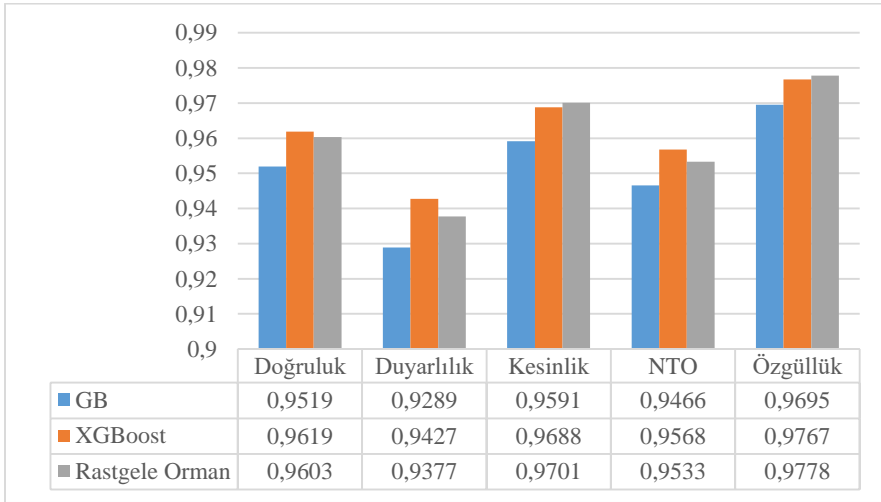
RO'da sınıflandırma işlemi, orijinal eğitim veri setinden rassal örnekler seçilmesi yoluyla rassal veri setleri üretir. Oluşturulan veri setinde eğitilen bir karar ağacıyla bir orman oluşturur. Ormanın nihai kararı, karar ağaçlarının oylamasıyla gerçekleşir (Ecer vd., 2018: 6). Bir dizi

zayıf sınıflandırıcının performansının tek sınıflandırıcıdan genellikle daha iyi olduğu kabul edilmektedir. Bu durum, model sonucuna daha yüksek bir kesinlik ve genelleme yeteneği kazandırır (Oshiro vd., 2012: 154; Zhang vd., 2020: 9).

5. BULGULAR

Çalışmanın modelleme aşamasında Tablo 1’de detayları verilen 21 değişken girdi değişkeni olarak kullanılmış olup, havayolu ulaşımından tatmin olup olmama değişkeni ise çıktı değişkeni olarak tanımlanmıştır. Topluluk öğrenmesi yöntemleriyle kurulan modellere ait en iyi sonuçlar ve bu sonuçlardan elde edilen değerlendirme kriterleri sonuçları elde edilerek modellerin başarımları yorumlanmıştır. Veri setinde yer alan gözlemlerin %70’i Eğitim %30’u ise Test Verisi olarak ayrılmıştır. Modellere uygulanan çapraz doğrulama (k-fold cross validation) sayısı 5’dir. Kullanılan modellere ait parametreler; Rastgele Orman: Bölünme Kriteri; Gini İndeksi, Ağaç Derinliği: 3, Model Sayısı: 100. XGBoost’a ait parametreler; η -Eta: 0,3, α -Alfa: 1, λ -Lamda: 1, γ -Gama: 1, Maksimum Delta Adımı: 0, Maksimum Derinlik: 6. Gradient Boosting’a ait parametreler; Ağaç Derinliği: 5, Model Sayısı: 100, Öğrenme Hızı: 0,1’dir.

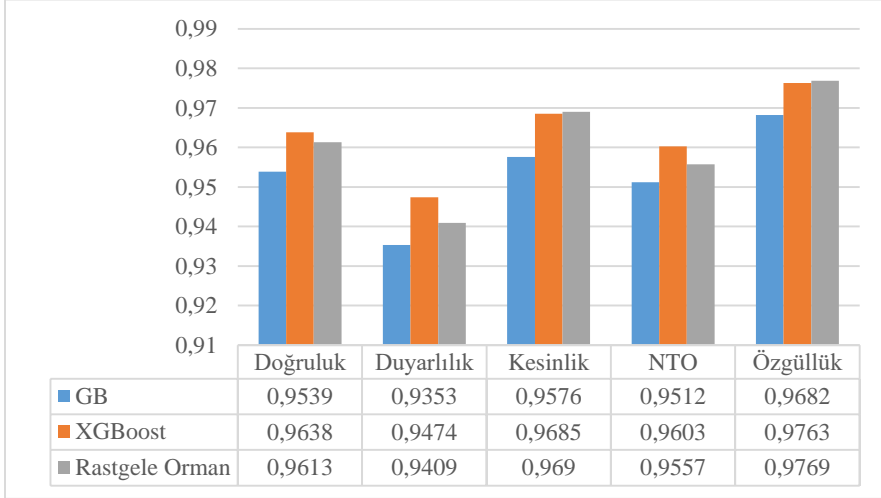
Çalışmada kullanılan veri seti dengeli bir veri setidir, daha açık bir ifade ile gerçekleştirdiği uçuştan memnun olan yolcu sayısı ile memnun olmayan yada nötr yolcu sayısı birbirine yakındır. Bu nedenle, uçuş seyahatinden nötr ya da memnun olmayan bireylerin tahmini için kurulan makine öğrenmesi modelleri sonuçlarına göre modellerin değerlendirilmesinde Doğruluk ölçütü önemlidir. Eğitim veri seti ile kurulan makine öğrenmesi modellerine ilişkin sonuçlar Şekil 2’de, Test veri seti ile kurulan modellere ait sonuçlar Şekil 3’te gösterilmiştir. Hem eğitim hem de test veri setinde elde edilen sonuçlar birbirine yakındır. Bu nedenle, modellerin tutarlı ve kararlı oldukları söylenebilir.



Şekil 2. Eğitim Veri Setine Ait Model Sonuçları

Test veri setinde; GB modelinde %95,39, XGBoost modelinde %96,38, RO ile kurulan modelde de %96,13 doğru tahmin başarısı görülmüştür. Problemin tanımı gereği, negatif sınıfların yani uçuş yolculuğundan nötr ya da memnun olmayanların tahmin edilmesi oldukça

önemlidir. Duyarlık ve Kesinlik ölçütleri pozitif sınıf tahminlerinin, Özgüllük ve Negatif Tahmin Oranı (NTO) ise negatif sınıf tahminlerinin yorumlanmasında önemlidir. Formül gereği Özgüllük ölçütü; toplam negatif gözlemler içerisinde doğru tahmin edilen gözlemlerin oranını verirken, NTO ölçütü ise toplam negatif tahminler içerisindeki doğru negatif tahminlerin oranını vermektedir.



Şekil 3. Test Veri Setine Ait Model Sonuçları

Modelleme sonuçlarına göre en iyi Duyarlık değeri XGBoost yöntemi ile en iyi Kesinlik değeri ise Rastgele Orman yöntemi ile elde edilmiştir. Duyarlılık ölçütü bakımından incelendiğinde ise uçuş yolculuğundan memnun olanlar içinde pozitif tahmin oranının en yüksek, yaklaşık %95 olduğu görülmektedir. Bu durumda gerçek memnun olanların modeller tarafından yakalanabilme oranının, Kesinlik değerine göre düşük olduğu, ancak toplam memnun olan yolcuların yaklaşık %95'inin doğru tahmin edilebildiği söylenebilir. Kesinlik ölçütü bakımından incelendiğinde uçuş yolculuğundan memnun olan şekilde tahmin edilen gözlemlerin içerisinde yaklaşık %97'si doğru tahmin edilmiştir. Buna göre Kesinlik bakımından modellerin başarımının yüksek olduğu söylenebilir.

Özgüllük ölçütü, uçuş yolculuğundan nötr ya da memnun olmayan tüm yolcuların, tahmin sonucunda da nötr ya da memnun olmayan olarak bulunması açısından önemlidir. En iyi Özgüllük değeri Rastgele Orman yöntemi ile elde edilmiştir. Bu duruma göre nötr ya da memnun olmayan uçuşların yaklaşık olarak %98 civarında doğru tahmin edildiği görülmektedir. Negatif Tahmin Oranında ise nötr ya da memnun olmayan tahminlerin içerisinde doğru olan tahminlerin oranı en yüksek %96 civarında olduğu görülmektedir.

Bu bulgulara göre negatif değerlerin, yani uçuştan memnun olmayan veya nötr olan yolcuların tespit edilebilmesinde en uygun model Rastgele Orman yöntemi ile elde edilmiştir. Uçuşundan memnun değil veya nötr olarak tahmin edilen yolcuların yüksek oranda doğrulukla tahmin edilebilmesinde ise en uygun model XGBoost yöntemi ile elde edilmiştir.

6. SONUÇ ve TARTIŞMA

Gerçekleştirilen çalışmanın amacı uçuş seyahatinden nötr ya da memnun olmayan müşterilerin topluluk öğrenmesi yöntemleri ile yüksek doğrulukla tahmin edilmesidir. Araştırma kapsamında Gradient Boosting, XGBoost ve Rastgele Orman yöntemleri kullanılarak çalışma modellenmiştir. Kurulan modellerin sonuçlarına göre hesaplanan performans ölçütleri ile modellerin karşılaştırmalı olarak tahmin performansları kıyaslanmıştır.

Model başarımlarına göre eğitim ve test sonuçlarının birbirine yakındır, bu durum modellerin tutarlı ve kararlı olduğunu göstermektedir. Modelleme sonuçlarına göre kurulan modeller hem pozitif tahminlerde hem de negatif tahminlerde başarılı olmasına karşın, uçuşundan nötr ya da memnun olmayan gözlemlerin tahmin başarımları uçuşundan memnun olanların tahminlerine kıyasla biraz daha yüksektir.

Noviantoro ve Juang (2022) havayolu yolcularının memnuniyetini %95,42 doğruluk (Derin Öğrenme) ile tahmin ederken, Kesinlik, Duyarlılık, F ve AUC ölçütlerinde en iyi sonuçları Derin Öğrenme yöntemi ile kurulan modelde elde etmiştir. Gao vd. (2021) havayolu seyahat memnuniyetini %95,92 doğruluk (Rastgele Orman) ile tahmin ederken Hwang vd. (2020) ise müşterilerin aynı havayolunu tekrar kullanmalarını %83,42 doğruluk (XGBoost) değeri ile tahmin etmiştir. Gerçekleştirilen bu çalışmada XGBoost ile kurulan model en iyi Doğruluk değerine sahiptir ve model %96,38 ile doğru tahmin yapmaktadır. Problemin tanımı gereği, negatif sınıfların yani uçuş yolculuğundan nötr ya da memnun olmayanların tahmin edilmesi daha önemlidir. En yüksek Negatif Tahmin Oranı değeri XGBoost ile kurulan modelde elde edilmiştir ve yaklaşık olarak 0,96 olarak görülmektedir. En yüksek Özgüllük değeri ise Rastgele Orman ile kurulan modelde elde edilmiştir ve yaklaşık olarak 0,98 olarak görülmektedir.

Model sonuçlarına göre elde edilen bulguların uçak seyahat hizmetlerindeki firmaların yöneticilerinin planlama yapma ve rasyonel karar almalarında bir fikir oluşturabileceği düşünülmektedir. Çalışmanın verileri Amerika Birleşik Devletleri'ndeki uçuş firmalarından derlenen verilerden oluşturulmuştur, bu nedenle gerçekleştirilen çalışma veri setinin derlendiği kapsam ile sınırlıdır. Başka ülkelerden ve/veya Türkiye'den derlenen verilerle benzer çalışmalar yapılabilir. Müşterilerin memnun olma ya da olmama durumlarını etkileyen faktörlerin tahmini gerçekleştirilebilir, diğer makine öğrenmesi yöntemleri kullanılarak model başarımlarının kıyaslamaları yapılabilir.

KAYNAKÇA

- ABDI, A. M. (2020). "Land Cover And Land Use Classification Performance of Machine Learning Algorithms In A Boreal Landscape Using Sentinel-2 Data", *GIScience & Remote Sensing*, 57(1): 1-20.
- AGARWAL, I., & GOWDA, K. R. (2021). "The Effect of Airline Service Quality on Customer Satisfaction and Loyalty in India". *Materials Today: Proceedings*, 37, Part 2: 1341-1348.
- AL DAOUD, E. (2019). "Comparison Between Xgboost, Lightgbm and Catboost Using A Home Credit Dataset", *International Journal of Computer and Information Engineering*, 13(1): 6-10.

- ALPAYDIN, E. (2010). *Introduction To Machine Learning* (second edition), Cambridge, Massachusetts, London, England, The MIT Press.
- AN, M., & NOH, Y. (2009). "Airline Customer Satisfaction and Loyalty: Impact of In-Flight Service Quality", *Service Business*, 3: 293-307.
- ANGGRAINA, A., PRIMARTHA, R., & WIJAYA, A. (2019). "The Combination of Logistic Regression and Gradient Boost Tree For Email Spam Detection", *Journal of Physics: Conference Series* (Vol. 1196, 012013), 1-6.
- BAYDOGAN, C., & ALATAS, B. (2019). "Detection of Customer Satisfaction on Unbalanced and Multi-Class Data Using Machine Learning Algorithms", In *2019 1st International Informatics and Software Engineering Conference (UBMYK)* (1-5).
- BENTÉJAC, C., CSÖRGŐ, A., & Martínez-Muñoz, G. (2021). "A Comparative Analysis of Gradient Boosting Algorithms". *Artificial Intelligence Review*, 54: 1937-1967.
- BREIMAN, L. (2001). "Random Forests". *Machine learning*, 45(1): 5-32.
- BUSTILLO, A., DÍEZ-PASTOR, J. F., QUINTANA, G., & GARCÍA-OSORIO, C. (2011). Avoiding neural network fine tuning by using ensemble learning: application to ball-end milling operations. *The International Journal of Advanced Manufacturing Technology*, 57: 521-532.
- CHEN, T., & GUESTRIN, C. (2016). "Xgboost: A Scalable Tree Boosting System", In *KDD '16: Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining* (785-794).
- ÇELİKKOL, E. S., UÇKUN, C. G., TEKİN, V. N., & ÇELİKKOL, Ş. (2012). "Türkiye'de İç Hatlardaki Havayolu Taşımacılığında Müşteri Tercihi ve Memnuniyetini Etkileyen Faktörlere Yönelik Bir Araştırma", *İşletme Araştırmaları Dergisi*, 4(3): 70-81.
- DPT, DEVLET PLANLAMA TEŞKİLATI (2001). "Sekizinci Beş Yıllık Kalkınma Planı, Ulaştırma Özel İhtisas Komisyonu Raporu Hava Yolu Ulaştırma Alt Komisyon Raporu", DPT:2584, ÖİK:596, Ankara.
- DUMAN, S., ELEWI, A., & YETGİN, Z. (2022). "Distance Estimation From a Monocular Camera Using Face and Body Features", *Arabian Journal for Science and Engineering*, 47: 1547-1557.
- ECER, O., YETGİN, Z., & CELİK, T. (2018). "Air Write Letter Recognition Using Random Forest Classification on Arduino Dataset", *International Journal of Scientific and Technological Research*, 4(7): 1-9.
- ERCAN, U. (2021). "İnternette Alışveriş Yapan Hanelerin Rastgele Orman Yöntemiyle Tahmin Edilmesi", *Kafkas Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 12(24): 728-752.
- FLORES, V., & KEITH, B. (2019). "Gradient Boosted Trees Predictive Models for Surface Roughness in High-Speed Milling in the Steel and Aluminum Metalworking Industry", *Complexity*, 1536716, 1-15.

- FRIEDMAN, J. H. (2001). "Greedy Function Approximation: A Gradient Boosting Machine", *Annals of Statistics*, 29(5): 1189-1232.
- GIAO, H. N. K. (2017). "Customer Satisfaction of Vietnam Airline Domestic Service Quality", *International Journal of Quality Innovation*. 3(10): 1-11.
- GAO, K., YANG, Y., & QU, X. (2021). "Examining Nonlinear and Interaction Effects of Multiple Determinants on Airline Travel Satisfaction. *Transportation Research Part D: Transport and Environment*, 97, 102957: 1-23.
- HUSSAIN, R., AL NASSER, A., & HUSSAIN, Y. K. (2015). "Service Quality and Customer Satisfaction of A UAE-Based Airline: An empirical investigation", *Journal of Air Transport Management*, 42: 167-175.
- HWANG, S., KIM, J., PARK, E., & KWON, S. J. (2020). "Who Will Be Your Next Customer: A Machine Learning Approach To Customer Return Visits In Airline Services", *Journal of Business Research*, 121: 121-126.
- JIANG, H., & ZHANG, Y. (2016). "An Investigation of Service Quality, Customer Satisfaction and Loyalty in China's Airline Market", *Journal of Air Transport Management*, 57: 80-88.
- KAGGLE (2022). *Airline Passenger Satisfaction*, <https://www.kaggle.com/datasets/teejmahal20/airline-passenger-satisfaction>.
- KHAN, M. Y., QAYOOM, A., NIZAMI, M. S., SIDDIQUI, M. S., WASI, S., & RAAZI, S. M. K. U. R. (2021). "Automated Prediction of Good Dictionary EXamples (GDEX): A Comprehensive Experiment with Distant Supervision, Machine Learning, and Word Embedding-Based Deep Learning Techniques". *Complexity*, vol. 2021, 2553199, 1-18.
- KUMAR, G., YADAV, S. S., YOGITA & PAL, V. (2022). "Machine Learning Based Framework to Predict Finger Movement for Prosthetic Hand", *IEEE Sensors Letters*, 6(6): 1-4.
- KUMAR, Y. J. N., SPANDANA, V., VAISHNAVI, V. S., NEHA, K., & DEVI, V. G. R. R. (2020). "Supervised Machine Learning Approach for Crop Yield Prediction in Agriculture Sector", In *2020 5th International Conference on Communication and Electronics Systems (ICCES) (736-741)*. IEEE.
- KUMAR, S., & ZYMBLER, M. (2019). "A Machine Learning Approach To Analyze Customer Satisfaction From Airline Tweets", *Journal of Big Data*, 6:62, 1-16.
- KÜÇÜK ÇIRPIN, B., & KURT, D. (2016). "Havayolu Taşımacılığında Hizmet Kalitesi Ölçümü", *Journal of Transportation and Logistics*, 1(1): 83-98.
- MARIESCU-ISTODOR, R., & JORMANAINEN, I. (2019). "Machine Learning for High School Students", In *Koli Calling '19: Proceedings of the 19th Koli Calling International Conference On Computing Education Research*, 10, (1-9).

- NOVIANTORO, T., & HUANG, J. P. (2022). "Investigating Airline Passenger Satisfaction: Data Mining Method", *Research in Transportation Business & Management*, 43, 100726: 1-13.
- OSHIRO, T. M., PEREZ, P. S., & BARANAUSKAS, J. A. (2012). "How many trees in a random forest?", *International Workshop on Machine Learning and Data Mining in Pattern Recognition* in (154-168). Springer-Verlag, Berlin, Heidelberg
- PAL, M. (2005). "Random Forest Classifier For Remote Sensing Classification", *International Journal of Remote Sensing*, 26(1): 217-222.
- PONRAJ, A. S., & VIGNESWARAN, T. (2020). "Daily Evapotranspiration Prediction Using Gradient Boost Regression Model for Irrigation Planning", *The Journal of Supercomputing*, 76: 5732-5744.
- RAJULA, H. S. R., VERLATO, G., MANCHIA, M., ANTONUCCI, N., & FANOS, V. (2020). "Comparison of Conventional Statistical Methods with Machine Learning in Medicine: Diagnosis, Drug Development, and Treatment", *Medicina*, 56, 455: 1-10.
- RUMORA, L., MILER, M., & MEDAK, D. (2020). "Impact of Various Atmospheric Corrections on Sentinel-2 Land Cover Classification Accuracy Using Machine Learning Classifiers", *ISPRS International Journal of Geo-Information*, 9, 277: 1-23.
- TOPAL, B., ŞAHİN, H., & TOPAL, B. (2019). "Havayolu İle Yolcu Taşımacılığında Müşteri Memnuniyetini Etkileyen Faktörlerin Belirlenmesi: İstanbul Hava Limanları Örneği", *Balkan Sosyal Bilimler Dergisi*, 8(16): 119-128.
- ÜSTÜNER, M., ABDİKAN, S., BİLGİN, G., & ŞANLI, F. B. (2020). "Hafif Gradyan Artırma Makineleri ile Tarımsal Ürünlerin Sınıflandırılması", *Türk Uzaktan Algılama ve CBS Dergisi*, 1(2): 97-105.
- VINOD, B. (2021). "Artificial Intelligence in Travel", *Journal of Revenue and Pricing Management*, 20: 368-375.
- VAN ASSCHE, A. & BLOCKEEL, H. (2007). Seeing the forest through the trees: Learning a comprehensible model from an ensemble. Kok, J. N., Koronacki, J., Mantaras, R. L. d., Matwin, S., Mladenič, D., Skowron, A. (Ed) in *Machine Learning: ECML 2007. ECML 2007. Lecture Notes in Computer Science()*, vol 4701, (418-429). Springer, Berlin, Heidelberg.
- ZHANG, L., LIU, Z., REN, T., LIU, D., MA, Z., TONG, L. & LI, S. (2020). "Identification of Seed Maize Fields with High Spatial Resolution and Multiple Spectral Remote Sensing Using Random Forest Classifier", *Remote Sensing*, 12, 362: 1-19.