



**TURİZM ENDÜSTRİSİNDE OTEL REZERVASYON İPTALLERİNİN MAKİNE ÖĞRENMESİ
YÖNTEMLERİ İLE TAHMİNİ (HOTEL BOOKINGS CANCELLATION PREDICTION WITH
MACHINE LEARNING METHODS IN TOURISM INDUSTRY)****

Uğur ERCAN¹ (orcid.org/ 0000-0002-9977-2718)

Sezgin IRMAK^{2*} (orcid.org/ 0000-0003-2188-7344)

¹Akdeniz Üniversitesi, Enformatik Bölümü, Antalya, Türkiye

²Akdeniz Üniversitesi, Uygulamalı Bilimler Fakültesi, Yönetim Bilişim Sistemleri Bölümü, Antalya, Türkiye

Özet

Turizm endüstrisinde otel rezervasyonlarının iptallerinin tahmin edilebilmesi, otellerin gelir yönetimleri açısından önemli olarak nitelendirilebilir. Bu çalışmada 119 bin 390 rezervasyon verisinin kayıtlı olduğu veri seti kullanılarak rezervasyon iptallerini tahmin eden makine öğrenmesi modelleri geliştirilmeye çalışılmıştır. Yapay Sinir Ağları, Destek Vektör Makineleri, Sınıflama ve Regresyon Ağaçları ile Rastgele Orman olmak üzere dört farklı makine öğrenmesi yöntemi kullanılmıştır. Geliştirilen modeller birbirleri ile kıyaslamalı olarak incelenmiştir. En yüksek doğruluk oranı %84,97 ile Destek Vektör Makineleri modeli ile elde edilmiş olup modeller ayrıca Kesinlik, Duyarlılık, Özgüllük ve Negatif Tahmin Edicilik Değeri ölçütleri bakımından da değerlendirilmiştir. Modellerden elde edilen yüksek doğruluk ve hassasiyet değerleri, makine öğrenmesi yöntemlerinin otel rezervasyonlarının kesinleştirilmesi veya iptali gibi tahmin işlemlerinde kullanılabileceğini göstermektedir. Bu sayede rezervasyon iptal olasılıkları yüksek misafirler için planlama yapma imkânı sağlanabilecektir.

Anahtar Kelimeler: Turizm Endüstrisi, Otel Rezervasyon İptalleri, Makine Öğrenmesi

Abstract

Predicting the cancellations of hotel bookings could be considered as important for revenue management of hotels in tourism industry. In this study, machine learning models were built to predict bookings cancellations using dataset which consists of 119,390 booking data. Four different machine learning methods were used including Artificial Neural Networks, Support Vector Machines, Classification and Regression Trees, and Random Forest. The built models were evaluated comparatively. The highest accuracy ratio, which is 84.97%, was obtained by Support Vector Machines model, additionally models were evaluated by using Precision, Recall (Sensitivity), Specificity, and Negative Predictive Value metrics. Models have high accuracy and sensitivity values that machine learning methods can be used for prediction processes of booking confirmations or cancellations. Thus, planning opportunity could be acquired for guests who have higher cancellation probabilities.

Key Words: Tourism Industry, Hotel Bookings Cancellations, Machine Learning

** Bu makalenin ön çalışması, 2021 yılı Kasım ayında III. Uluslararası Sürdürülebilir Turizm Kongresinde sözlü bildiri olarak sunulmuş ve genişletilmiş özet olarak basılmıştır.

* Sorumlu yazar: sezgin@akdeniz.edu.tr

DOI: 10.33083/joghat.2022.112

Giriş

Turizm endüstrisi, ekonomik kalkınma açısından, diğer faktörlerin yanı sıra turistler tarafından mal ve hizmetlerin tüketimi, vergiler, işletmelerin gelişimi ve istihdam fırsatları yoluyla ülkeler için gelir yaratır. Fakat turizmin etkinliğinin gerçekleştirildiği ortam onu sürekli bir belirsizlikle karşı karşıya kalmaya zorlamaktadır (Sánchez-Medina ve C-Sánchez, 2020). Bu anlamda turizm sektörü, politik istikrarsızlık, terör saldırıları, hava durumu, doğal afetler, yangınlar, kazalar, ani hastalıklar, program çakışmaları ve misafirlerin öngörülemez aile yükümlülükleri gibi birçok faktörden etkilenmektedir (Chen ve Xie, 2013; Falk ve Vieru, 2018). Konaklama endüstrisinde müşteriler rezervasyon yaparken ücretli iptal, ücretsiz iptal ya da ayrıcalıklı (premium) bir fiyatla rezervasyon yapma seçeneğine sahiptir (Masiero vd., 2020). Otel rezervasyonunu iptal etme seçeneği kendileri için faydalı olsa da otel yöneticilerinin tercih etmediği bir durumdur (Falk ve Vieru, 2018). Rezervasyon iptallerinin, talep yönetimi kararları üzerinde önemli bir etkisi vardır ve gelir yönetimi performansı açısından kritik bir araç olan doğru tahminlerin üretimini sınırlarlar. Otel yöneticileri hangi rezervasyonların iptal edilebileceğini belirleyerek hizmet, oda yükseltme, indirim, gösteri/eğlence parklarına giriş veya diğer avantajlar gibi olası iptalleri engellemek için önlemler alabilir, ya da iptallerin neden olduğu sıkıntıları azaltmak için oteller katı iptal politikaları ve çifte rezervasyon stratejileri uygulayabilirler (Antonio vd., 2017a). Bununla beraber seyahat işletmelerinin, iptaller için müşterilerden ücret alması durumunda, iptal politikasının gelir yönetimi optimizasyonu için bir araç olabileceği belirtilmiştir (Chen ve Xie, 2013).

Falk ve Vieru (2018) gerçekleştirdikleri çalışmada otellerde genel iptal oranının %8 olduğunu belirtirken, oteller aldığı rezervasyonların %20 ila %60'ını etkileyen iptallerle karşı karşıya kalabilmektedir (Antonio vd., 2019a). Otellerin maksimum kâr elde etmeleri için sınırlı sayıdaki odalarını doğru zamanda doğru misafire tahsis etmesi gerekir (Boz vd., 2018). Tahmin, yönetimin karar verme faaliyetinin ayrılmaz bir parçasıdır, bu nedenle gelecekteki talebin iyi bir tahminine sahip olmak, gelir yönetimi analistlerinin envanterleri hakkında iyi kararlar almasına yardımcı olur (Nikseresht ve Ziarati, 2015). Morales ve Wang (2010) bir rezervasyonun iptal edilip edilmeyeceğini tahmin etmenin zor olduğunu belirtmiştir. Talep tahmini, başarılı bir gelir yönetimi sisteminde kritik bir rol oynar (Nikseresht ve Ziarati, 2015) bununla beraber talebin iptali de gelir yönetimini etkileyebilir. Rezervasyonun iptal edilip edilmeyeceğini önceden tahmin etmek, otelin stratejik ve operasyonel kararlarını almada önemli bir unsur olarak değerlendirilebilir. Müşterilerin otel ve seyahat rezervasyonlarını iptal etmek için yönlendiren sebepler hakkında çok az şey bilindiği gibi rezervasyon iptalinin nasıl önlenebileceği hakkında da detaylı bir bilgi bulunmamaktadır. Spesifik olarak, farklı önleme yaklaşımlarının etkinliğinin krizler ve turist bölümleri (segmentleri) arasında farklılık gösterip göstermediği belirsizdir (Hajibaba vd., 2016). İptal oranının tahmin edilmesi yalnızca fazla rezervasyon düzeylerini belirlemek için değil aynı zamanda net talebin tahminine de katkıda bulunmaktadır (Morales ve Wang, 2010). Bu çalışmada otel rezervasyon iptallerinin tahmin edilmesinde makine öğrenmesi yöntemlerinin kullanılabilirliğinin araştırılması ve belirli yöntemlerin tahmin etmedeki başarımlarının karşılaştırılması olarak incelenmesi amaçlanmaktadır.

Kavramsal Çerçeve

Rezervasyon iptalleri hali hazırda hizmet sektörlerinde, özellikle de konaklama sektöründe gelir yönetimi açısından incelenmekte olan bir konudur (Antonio vd., 2017b). Literatürde, Morales ve Wang (2010) veri madenciliği ile otellerin iptal oranını tahmin ederken, Chen vd. (2011) ise iptal politikalarını ve bu politikaların turistlerin fırsat arama davranışlarını şekillendirmedeki rolünü incelemiştir. Chen ve Xie (2013) otellerin iptal tipolojisini belirlemiş, Caicedo-Torres ve Payares (2016) ise otellerin günlük doluluk oranlarını makine öğrenmesi yöntemleri ile tahmin etmişlerdir. Elde edilen sonuçlara göre Ridge regresyon modeli, test veri setinde %8,66 MAPE değeri ile diğer modellerden daha iyi performans göstermiştir. Antonio vd. (2017a) otellerin rezervasyonlarının iptal durumlarını makine öğrenmesi yöntemleri ile tahmin etmiştir. Tatil ve şehir otelleri için %84 doğruluk elde edilirken, tatil otellerinde kesin rezervasyonların %73,6'sını şehir otellerinde ise %71,2'sini doğru tahmin etmiştir. Benzer şekilde Boz vd. (2018) otellerin rezervasyonlarının iptal durumlarını makine öğrenmesi yöntemleri ile tahmin ederken modelleme aşamasında en iyi sonucu %73 doğruluk değeri ile C4.5 yöntemi ile kurulan model vermiştir. Falk ve Vieru (2018) otel rezervasyonları ile ilgili olarak iptal davranışını etkileyen faktörleri incelemiştir. Andriawan vd. (2020) makine öğrenmesi yöntemlerini kullanarak otellerin rezervasyonlarının iptal durumlarını tahmin etmiştir. En iyi tahmin sonucu RO modeli ve %87,25 doğruluk ile elde edilmiştir. Saputro ve Nanang (2021) otellerin rezervasyon iptallerinin sayısındaki artışı azaltmak ve iptali tahmin etmek için makine öğrenmesi yöntemlerini kullanmışlardır. RO ve Extra Tree Classifier yöntemleri ile kurulan modellerin hassasiyet oranlarının yüksek olduğu, RO modeli ile kurulan modelin hassasiyet (recall) değerinin yüksek olduğu daha açık bir ifade gerçek pozitifleri %79 ile

doğru tahmin etmiştir. Bu sonuçlar, rezervasyon iptallerini yüksek doğrulukla tahmin etmenin mümkün olduğunu kanıtlamaktadır.

İncelenen çalışmalarda XGBoost, YSA, Karar Ağaçları, Rastgele Orman, Extra Tree Classifier, Minimum Squared Expected Error tree, Destek Vektör Makineleri ve k-En Yakın Komşu makine öğrenmesi yöntemleri ile Lojistik Regresyon, Kernel Lojistik Regresyon, Ridge Regresyon, Kernel Ridge Regresyon ekonometrik modelleri kullanılmıştır.

Bu araştırma kapsamında Destek Vektör Makineleri (DVM), Yapay Sinir Ağları (YSA), Sınıflama ve Regresyon Ağaçları (C&RT) ile Rastgele Orman (RO) yöntemleri kullanılarak otellerdeki rezervasyon iptallerinin tahmin edilmesi süreçleri izlenmiştir.

Veri Seti

Araştırmada Hotel Booking Demand Datasets (Antonio vd., 2019b) veri seti olarak kullanılmıştır. Veri setinde Tatil (Resort) Oteli (40.060 gözlem) ve Şehir (City) Oteli (79.330 gözlem) olmak üzere iki ayrı veri setinde toplam 119.390 gözlem yer almaktadır. Her bir gözlem, bir rezervasyon kaydını temsil etmekte olup, 44.224 adet rezervasyon iptal edilmiş olarak kayıtlıdır. Veri setinde yer alan değişkenlere ilişkin açıklamalar Tablo 1’de verilmiştir. Çalışmada kullanılan veriler, kamuya ve araştırmacılara açık kaynaklardan sağlanan ikincil veri olup etik kurul izni gerektirmemektedir.

Tablo 1. Değişken Tanımları

Değişken Adı	Tipi	Açıklama
HotelType	Kategorik	Tatil Oteli veya Şehir Oteli olmak üzere 2 kategori.
LeadTime	Sayısal	Rezervasyonun PMS’e girilmesinden varış tarihine kadar geçen gün sayısı.
ArrivalDateMonth	Kategorik	Varış Tarihi Ayı. Ocak-Aralık arası 12 ay.
StaysInWeekendNights	Sayısal	Konaklanan veya rezerve edilen hafta sonu gece sayısı (Cumartesi, Pazar).
StaysInWeekNights	Sayısal	Konaklanan veya rezerve edilen hafta içi gece sayısı.
Adults	Sayısal	Konaklayan yetişkin sayısı.
Children	Sayısal	Konaklayan çocuk sayısı.
Babies	Sayısal	Konaklayan bebek sayısı.
Meal	Kategorik	Tanımsız/Yemek yok, Yatak-Kahvaltı, Yarım Pansiyon ve Tam Pansiyon olmak üzere alınan yemek paketi.
Country	Kategorik	Rezervasyonun kaynak ülkesi.
MarketSegment	Kategorik	Direkt, Kurumsal, Grup, Havayolu, Hediye, Online Seyehat Acentası, Offline Seyehat Acentası/Tur Operatörü ve Tanımsız olmak üzere pazar bölümü.
DistributionChannel	Kategorik	Direkt, Kurumsal, GDS, Seyehat Acentası/Tur Operatörü ve Tanımsız olmak üzere rezervasyon kanalı.
IsRepeatedGuest	Kategorik	Tekrarlanan misafir (1) veya değil (0)
PreviousCancellations	Sayısal	Mevcut rezervasyona kadar müşterinin iptal etmiş olduğu rezervasyon sayısı.
PreviousBookingsNotCancelled	Sayısal	Mevcut rezervasyona kadar müşterinin iptal etmemiş olduğu rezervasyon sayısı.
BookingChanges	Sayısal	Rezervasyonun PMS’e girilmesinden, otele giriş veya rezervasyon iptaline kadar geçen sürede kaç defa değişiklik veya düzenleme yapıldığı.
DepositType	Kategorik	Depozitosuz, Geri Ödemesiz ve Geri Ödemeli olmak üzere alınan ödeme türü.

Değişken Adı	Tipi	Açıklama
CustomerType	Kategorik	Kontrat, Grup, Geçici, Geçici-parti olmak üzere 4 kategoride rezervasyon tipi.
ADR	Sayısal	Ortalama Günlük Ücret
RequiredCarParkingSpaces	Sayısal	Müşteri tarafından istenilen araç park yeri adedi.
TotalOfSpecialRequests	Sayısal	Müşteri tarafından talep edilen toplam özel istek adedi.
IsCancelled*	Kategorik	Rezervasyon iptal edilmiş (1) veya iptal edilmemiş (0) olmak üzere rezervasyonun nihai durumu.

Kaynak: Antonio vd., 2019b., *Hedef değişken.

Yöntem

Makine öğrenmesi, veri ve bilgi arasındaki temel ilişkileri sentezlemek için sistematik olarak algoritmalar uygulayan bir yapay zekâ dalıdır. İnternette arama, reklam yerleştirme, kredi puanlama, borsa tahmini, gen dizilim analizi, davranış analizi, ilaç geliştirme, hava tahmini, büyük veri analitiği ve daha birçok uygulamada yaygın olarak kullanılmaktadır (Awad ve Khanna, 2015). Bu çalışmada, makine öğrenmesi yöntemleri turizm alanında kullanılarak Avrupa'daki otellerde konaklayan müşterilerin rezervasyonlarının iptal durumlarının tahmin edilmesine yönelik bir uygulama gerçekleştirilmiştir. Araştırmada makine öğrenmesi modellemesi için 4 farklı yöntem kullanılmıştır. Sırasıyla, Destek Vektör Makineleri (RBF Çekirdek Fonksiyonu), Yapay Sinir Ağları (İleri Beslemeli Ağ ve Geri Yayılım Algoritması), Sınıflama ve Regresyon Ağaçları (Classification & Regression Trees – C&RT) ve Rastgele Orman (Random Forest – RO) yöntemleri ile modelleme yapılmış olup sonuçları karşılaştırmalı olarak incelenmiştir. Model başarımlarının karşılaştırılmasında Karışıklık Matrisi (Confusion Matrix), Doğruluk (Accuracy), Kesinlik (Precision), Duyarlılık (Sensitivity), Özgüllük (Specificity) ve Negatif Tahmin Edebilirlik Değeri (Negative Predictive Value) kriterleri kullanılmıştır.

Yapay Sinir Ağları (YSA)

Yapay sinir ağları, biyolojik organizmalardaki öğrenme mekanizmasını simüle eden popüler makine öğrenme tekniklerinden birisidir (Aggarwal, 2018). Doğrusal veya değil, herhangi bir duruma ait girdi-çıkıtı arasındaki ilişkiyi, mevcut örneklerden öğrenerek daha önce karşılaşılmamış durumları, önceki deneyimlerini kullanarak olaya ilişkin çözümler üretebilmesi, algoritmik çözümü bulunamamış problemlerin çözülmesinde kullanımını artırmıştır (Yakut vd., 2014). Tıpkı biyolojik sinir ağlarında olduğu gibi YSA'larda yer alan nöronlar birbirine bağlanmak suretiyle bir ağ yapısı oluşturur. Bir sinir ağı, ağırlıklı bağlantılar aracılığıyla birbirine bağlanan nöronlar adı verilen bir dizi temel hesaplama biriminden oluşur (Giudici, 2003). Nöronlar aslında, doğrusal olmayan, parametrelili, sınırlı bir fonksiyondur (Dreyfus, 2005) ve otonom bir hesaplama birimini temsil eder. Nöronlar aktivasyonunu bir dizi sinyal olarak girdileri alır, aktivasyonun ardından bir çıkış sinyali üretir. Her giriş sinyali bir bağlantı ağırlığı ile ilişkilendirilir. Ağırlık, giriş sinyalinin nöron tarafından iletilen son çıktıyı üretmedeki göreceli önemini belirler (Giudici, 2003).

Destek Vektör Makineleri

DVM örüntü tanıma, sınıflandırma, zaman serisi tahmini ve regresyon analizi gibi birçok makine öğrenmesi görevi için problemlerin çözümünde kullanılan denetimli bir öğrenme modelidir (Sapankevych ve Sankar, 2009; Poongodi vd., 2020). İstatistiksel öğrenme teorisine dayanan, örüntü tanıma problemlerini çözmek için Vapnik tarafından geliştirilen DVM'nin temel amacı yüksek boyutlu bir öznelik uzayında sınıflar arasında en büyük marj ile en uygun hiperdüzlemi bularak verileri sınıflandırmaktır (Kim vd., 2013). Yapay Sinir Ağları deneysel risk minimizasyonu prensibini benimsemektedir. DVM'ler ise YSA gibi geleneksel öğrenme metodlarından farklı olarak yapısal risk minimizasyonu prensibini benimser. Böylelikle DVM'de, YSA modellerinde görülen aşırı uyum ve yerel minimuma takılma gibi tipik dezavantajları ortadan kaldırılmıştır (Lu, 2002). DVM'de girdi uzayı doğrusal olarak ayrılabiliriyorsa Doğrusal DVM, doğrusal olarak ayrılamiyorsa Doğrusal Olmayan DVM kullanılır. Doğrusal DVM yalnızca ayırt edilebilir doğrusal verilere uygulanabilen en basit DVM modelidir (Çevik ve Kayakuş, 2020). Doğrusal olarak ayrıştırılmayan veri yapılarında ise çekirdek fonksiyonları kullanılır. DVM, verileri yüksek boyutlu bir düzleme doğrusal olmayan bir eşleme işlemi yapar. Daha sonra, bu yeni uzayda basit doğrusal karar sınırları oluşturmak için basit doğrusal fonksiyonlar kullanılır. Burada önemli olan nokta uygun çekirdek fonksiyonunun seçilmesidir (Aydın Atasoy ve Tabak, 2018). Literatürde yaygın olarak kullanılan çekirdek fonksiyonları RBF, Sigmoid, Lineer ve Polinomial'dır (Çevik ve Kayakuş, 2020).

Sınıflandırma ve Regresyon Ağaçları (C&RT)

Sınıflandırma ve Regresyon Ağaçları (C&RT), 1984 yılında Breiman vd. (1984) tarafından geliştirilmiştir (Rokach ve Maimon, 2015). Teorisi ve matematiksel algoritması oldukça karmaşık olan C&RT, verilerin sınıflandırılması veya tahmin edilebilmesi için karar ağaçlarını kullanan gelişmiş bir araçtır (Steinberg ve Colla, 1995). Basit, parametrik olmayan, eksik ve aykırı değerlere sahip veri setlerini işleyebilen (Nisbet vd., 2018) sürekli ve kategorik değişkenlerin kullanılabilirdiği C&RT yönteminde, ağacın oluşturulması esnasında bölünme kriteri olarak Gini ve Least Squared Deviation ölçütleri kullanılmaktadır (Palmer vd., 2011). C&RT yöntemi, otomatik sınıf dengeleme, maliyet duyarlı öğrenme, dinamik özellik oluşturma ve olasılık ağacı tahmini sağlar (Steinberg, 2009). C&RT’de böl ve yönet prensibi kullanılmaktadır, ağacın yapısı, eksik verilerin ele alınma biçimi, bölünme ve budanma kriterleri temel farklılaşma özellikleridir (Kantardzic, 2019). Diğer karar ağacı modellerinin aksine her düğümde ikili bölünmeler oluşur (Rokach ve Maimon, 2015). Ağaç, herhangi bir durdurma kuralı kullanılmadan maksimum boyuta ulaşır. Esasen ağacın büyüme süreci veri eksikliğinden dolayı bölünme gerçekleşmediği için sonlanır. Maksimum boyutlu ağaç, Maliyet Karmaşıklığı Budama (Cost-Complexity Pruning) yöntemi ile köke doğru geri biçimde budanır. Budanacak bir sonraki bölme, ağacın eğitim verileri üzerindeki genel performansına en az katkıda bulunan bölmedir (Steinberg, 2009).

Rastgele Orman

Genellikle, aynı miktarda eğitim verisi verilen bir dizi zayıf sınıflandırıcının performansı, tek sınıflandırıcıdan daha iyidir. Yaygın olarak bilinen topluluk yöntemleri Bagging, Boosting ve Rastgele Orman yöntemleridir (Oshiro vd., 2012). Breiman (2001) tarafından geliştirilen bir topluluk öğrenme yöntemi olan RO hem sınıflandırma hem de regresyon problemlerinde kullanılmaktadır. Karar ormanındaki her bir ağaç orijinal veri setinden bootstrap örnekleme yapılarak oluşturulur. Karar verme aşamasında ormanda bulunan ağaçlardan elde edilen sınıf tahminleri bir araya getirilerek tek bir sınıf sonucu elde edilir. RO yöntemi, Bagging yöntemi ve Ho (1998) tarafından önerilen Rastgele Alt-uzay yöntemlerinin birleştirilmesinden oluşmaktadır (Uzbaş, 2017). Rastgele Orman yönteminin hızlı, eğitilmesi kolay (Wang vd., 2019), aykırı ve gürültülü değerlere karşı sağlam (Breiman, 2001), sürekli ve kesikli değerlerle çalışabilmesi gibi birçok avantajı vardır (Akın ve Terzi, 2020).

Bulgular

Araştırmanın makine öğrenmesi ile modelleme aşamasında Tablo 1’de verilen ilk 21 değişken girdi değişkenleri olarak kullanılmış olup, rezervasyonun iptal edilip edilmediğine ilişkin son değişken ise hedef değişken olarak tanımlanmıştır. Her bir modelleme yöntemi ile elde edilen en iyi model Yöntem kısmında verilmiş olan değerlendirme kriterleri kullanılarak model başarımları analiz edilmiştir. Gözlemlerin %70’i Eğitim Verisi Bölmesi (Training Partition), %30’u ise Test Verisi Bölmesi (Testing Partition) olarak ayrılmıştır. Modellerin tahmin başarımlarına ilişkin karşıtlık matrisleri Tablo 2’de verilmiştir.

Tablo 2. Eğitim ve Test Bölümleri İçin Karşıtlık Matrisleri

		Eğitim Veri Seti		Test Veri Seti	
		0	1	0	1
DVM	0	48.200	4.318	20.386	2.049
	1	8.236	22.763	3.722	9.467
YSA	0	47.689	4.815	20.342	2.076
	1	8.487	22.508	3.617	9.566
C&RT	0	48.214	4.447	20.544	1.961
	1	12.053	18.969	5.204	7.998
Rastgele Orman	0	47.480	5.181	20.232	2.273
	1	11.194	19.828	4.863	8.339

* Satırlar hedef değişkenin gerçek değerlerini, sütunlar ise tahmin değerlerini temsil etmektedir.

Otel rezervasyonlarının iptal tahmini için modelleme sonuçlarının değerlendirilmesinde toplam doğruluk oranları önemli bir ölçüt olmakla birlikte, özellikle 1 ile temsil edilen pozitiflerin yani iptal edilen rezervasyonların yüksek doğrulukla tahmin edilebilmesi de problemin niteliği bakımından oldukça önemlidir.

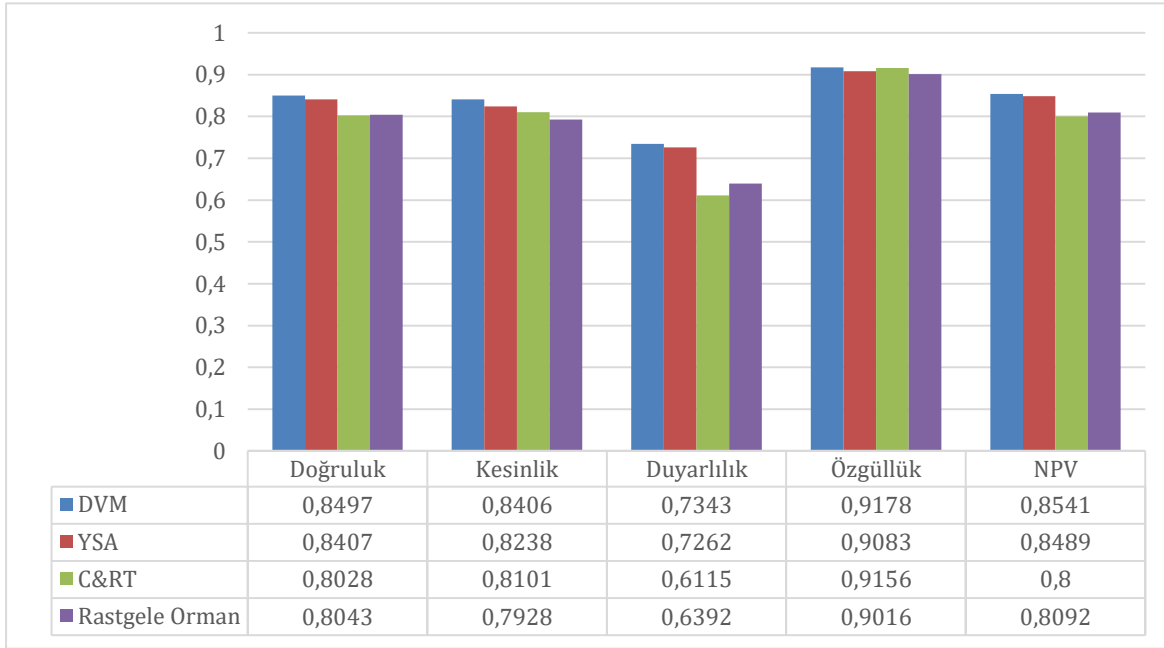
Karşılıklı matrisleri değerlendirildiğinde modellerin hem negatif (0) gözlemlerin hem de pozitif (1) gözlemlerin tahmin edilmesinde başarılı sonuçlar verdiği söylenebilir. Tahmin başarımları eğitim ve test bölmelerinde birbirine yakın sonuçlar verdiği için modellerin kararlı olduğu görülmektedir.

Model başarımlarının problemin niteliği gereği detaylı olarak ele alınabilmesi için Doğruluk (Accuracy), Kesinlik (Precision), Duyarlılık (Recall), Özgüllük (Specificity) ve Negatif Tahmin Edebilirlik Değeri (Negative Predictive Value - NPV) ölçütleri ile değerlendirilmesi uygun düşmektedir. Dört farklı modele eğitim aşamasına ilişkin değerlendirme sonuçları Şekil 1’de verilmiştir.

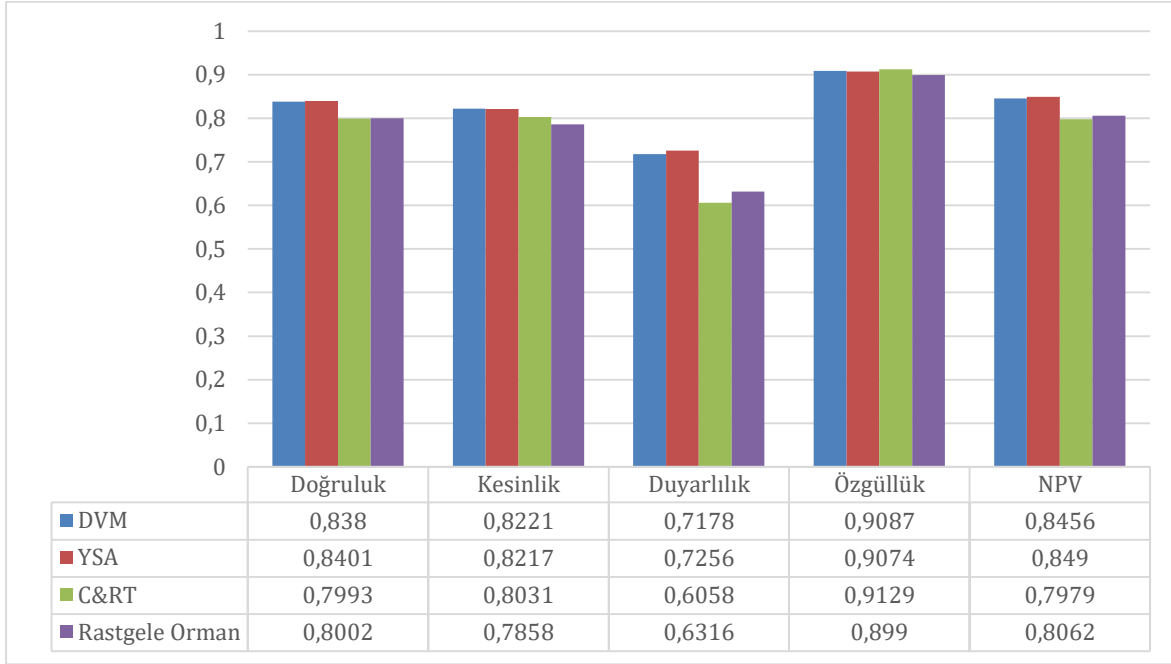
Modellerin performans ölçütlerine göre değerlendirilmesinde en yüksek doğruluk oranının DVM Eğitim Bölmesi için elde edildiği görülmektedir. Ancak Eğitim ve Test Bölmesi için birbirine yakın ve yüksek doğruluk oranları YSA modeli ile elde edilmiştir. Doğruluk ölçütü hedef değişken bakımından pozitif (1) ve negatif (0) tüm gözlemler için genel doğru tahmin oranını vermektedir.

Rezervasyon iptallerinin tahmininde pozitif yani iptal edilen/edilecek rezervasyonların tahmininin daha fazla öneme sahip olduğu söylenebilir. Bu durumda Kesinlik ve Duyarlılık ölçütlerinin incelenmesi gerekmektedir. Kesinlik ölçütü toplam pozitif tahminler içerisinde doğru pozitif tahminlerin oranını vermektedir. Duyarlılık ölçütü ise toplam pozitif gözlemler içerisinde doğru tahmin edilen gözlemlerin oranını vermektedir. Model sonuçlarına göre en yüksek kesinlik değeri DVM modelinde, daha tutarlı ve yüksek kesinlik değeri ise YSA modelinde görülmektedir. Eğitim ve test sonuçları birlikte değerlendirildiğinde iptal edilen/edilecek rezervasyon olarak tahmin edilen gözlemlerin içerisinde yaklaşık %82’si doğru tahmin edilmiştir. Bu da kesinlik bakımından modellerin başarımının yüksek olduğu şeklinde yorumlanabilir. C&RT ve Rastgele Orman modellerinde ise başarım oranları %81 ile %78 arasındadır.

Şekil 1. Eğitim Aşaması Performans Ölçütleri



Model değerlendirme sonuçlarına göre eğitim ve test aşaması ölçüt değerlerinin birbirine yakın sonuçlar verdiği görülmektedir. Bu durum eğitim aşamasında herhangi bir aşırı öğrenme sorunu olmadığına işaret ettiği gibi modellerin tutarlı tahminler yapabileceğinin de bir göstergesi olarak kabul edilebilir. Test aşamasına ilişkin model değerlendirme sonuçları Şekil 2’de sunulmuştur.

Şekil 2. Test Aşaması Performans Ölçütleri

Duyarlılık ölçütü bakımından incelendiğinde ise iptal edilen rezervasyonlar içinde pozitif tahmin oranının en yüksek yaklaşık %73 olduğu görülmektedir. Bu durumda gerçek iptallerin modeller tarafından yakalanabilme oranının kesinlik değerine göre düşük olduğu ancak toplam iptal edilenlerin 2/3'ünden fazlasının doğru tahmin edilebildiği söylenebilir. Özgüllük ölçütü negatif tahminler içinde doğru bir şekilde tahmin edilen negatiflerin oranını vermektedir. Bu durumda iptal edilmeyen rezervasyonların %90'lar dolayında doğru tahmin edildiği görülmektedir. Burada yüksek ve aynı zamanda kararlı Özgüllük değerlerini C&RT modelinin verdiği görülmektedir. Negatif Tahmin Edilebilirlik Değerinde (NPV) ise iptal edilmeyen rezervasyon tahminlerinin içerisinde doğru olan tahminlerin en yüksek yaklaşık %85'ler dolayında olduğu görülmektedir.

Sonuç ve Tartışma

Gerçekleştirilen çalışmanın amacı otellerdeki rezervasyon iptallerinin makine öğrenmesi yöntemleri ile tahmin edilmesidir. Araştırma kapsamında Destek Vektör Makineleri (DVM), Yapay Sinir Ağları (YSA), Sınıflama ve Regresyon Ağaçları (C&RT) ile Rastgele Orman (RO) yöntemleri kullanılarak otellerdeki rezervasyon iptallerini tahmin eden modeller elde edilmiştir. Modellemeler sonucu elde edilen karışıklık matrisleri ve hesaplanan performans ölçütleri ile modellerin tahmin performansları kıyaslanmıştır.

Genel olarak model sonuçlarının hem pozitif tahminlerde hem de negatif tahminlerde başarılı olduğu söylenebilir. Negatif gözlemlerin tahmin başarımları pozitif tahminlere göre bir miktar daha yüksektir. Ayrıca modellerin doğru tahmin oranlarının incelenmesinde eğitim ve test bölmelerinde başarımlarının birbirine yakın olması da modellerin kararlılıklarının göstergesi olarak yorumlanabilir.

Modeller kıyaslamalı olarak incelendiğinde Destek Vektör Makineleri ile elde edilen model küçük farklarla daha yüksek başarımlar göstermekte, ancak Yapay Sinir Ağları modeli ise daha kararlı bir yapıdadır. Sınıflama ve Regresyon Ağaçları (C&RT) modelinin ise Özgüllük ölçütü bakımından daha iyi sonuç verdiği görülmektedir. Rastgele Orman modeli ise çok küçük farklar ile daha düşük kalmıştır.

Çalışmanın problemi göz önüne alındığında, pozitif değerlerin yani iptal edilen rezervasyonların doğru tahmin edilmesinin, negatif gözlemlerin doğru tahmin edilmesine nazaran daha önemli olduğu söylenebilir. Bu durumda C&RT ve Rastgele Orman yöntemlerine göre Kesinlik ve Duyarlılık değerleri daha yüksek olan Destek Vektör Makineleri yöntemi ile Yapay Sinir Ağları yönteminin daha başarılı sonuç verdiği söylenebilir.

Antonio vd. (2017a) gerçekleştirdikleri çalışmada otellerin rezervasyonlarının iptal durumlarını %84 (XGBoost yöntemi) doğruluk ile tahmin ederken, Boz vd. (2018) %73 (C&RT yöntemi) doğruluk değeri ile tahmin etmiştir. Andriawan vd. (2020) ise otellerin rezervasyon iptallerini %87,25 (RO yöntemi) doğruluk ile tahmin etmiştir. Saputro ve Nanang (2021) ise rezervasyon iptallerinde gerçek pozitifleri (hassasiyet) %79 (RO yöntemi) ile doğru tahmin etmiştir. Mevcut çalışmada ise rezervasyon iptallerini tahmin etmek için dört farklı makine öğrenmesi yöntemi kullanılmıştır. YSA ile kurulan model, iptal edilen rezervasyonların yaklaşık olarak %84'ünü doğru bir şekilde tahmin etmiştir. C&RT yöntemi ile kurulan model iptal edilmeyen rezervasyonları ise yaklaşık olarak %91,3 ile doğru tahmin etmiştir.

Sonuç olarak bu çıkarımların otellerde karar alan ve planlama yapan yöneticiler için bir fikir oluşturabileceği ve stratejik karar vermelerine yardımcı olacağı düşünülmektedir. Otel rezervasyonlarının iptal durumlarının makine öğrenmesi ile tahmin edilmeye çalışıldığı bu çalışmada Avrupa'daki otellerden derlenen veriler kullanılmıştır. Bu yönüyle araştırma veri setinin derlendiği kapsam ile sınırlıdır. Türkiye'den veriler elde edilerek izlenen yöntem ile benzer çalışmalar yapılabilir. Ayrıca başka makine öğrenmesi yöntemleri uygulanarak bu kapsamda model başarımlarının farklı kıyaslamaları yapılabilir.

Not: Bu makale 12-14 Kasım 2021 tarihleri arasında çevrimiçi olarak düzenlenen III. Uluslararası Sürdürülebilir Turizm Kongresinde sözlü bildiri olarak sunulmuş ve genişletilmiş özet olarak basılmış "Turizm Endüstrisinde Otel Rezervasyon İptallerinin Makine Öğrenmesi Yöntemleri ile Tahmini" başlıklı bildirinin genişletilmiş halidir.

Kaynakça

- Aggarwal, C. C. (2018). *Neural networks and deep learning a textbook*, Cham, Switzerland: Springer International Publishing.
- Akın, P., ve Terzi, Y. (2020). Dengesiz veri setli sağkalım verilerinde cox regresyon ve rastgele orman yöntemlerin karşılaştırılması. *Veri Bilimi*, 3(1), 21-25.
- Andriawan, Z. A., Purnama, S. R., Darmawan, A. S., Ricko, Wibowo, A., Sugiharto, A. ve Wijayanto, F. (2020, Kasım). Prediction of hotel booking cancellation using CRISP-DM. 2020 4th International Conference on Informatics and Computational Sciences (ICICoS), Semarang, Indonesia.
- Antonio, N., de Almeida, A. ve Nunes, L. (2017a, Kasım). Predicting hotel bookings cancellation with a machine learning classification model. 2017 16th IEEE International Conference on Machine Learning and Applications (ICMLA), Cancun, Mexico.
- Antonio, N., De Almeida, A. ve Nunes, L. (2017b). Predicting hotel booking cancellations to decrease uncertainty and increase revenue. *Tourism & Management Studies*, 13(2), 25-39.
- Antonio, N., de Almeida, A. ve Nunes, L. (2019a). Big data in hotel revenue management: exploring cancellation drivers to gain insights into booking cancellation behavior. *Cornell Hospitality Quarterly*, 60(4), 298-319.
- Antonio, N., de Almeida, A. ve Nunes, L. (2019b). Hotel booking demand datasets, *Data in Brief*, 22, 41-49.
- Awad, M. ve Khanna, R. (2015). *Efficient Learning Machines: Theories, Concepts, and Applications For Engineers and System Designers*. New York: Apress.
- Aydın Atasoy, N. ve Tabak, D. (2018). Destek vektör makineleri kullanarak yüz tanıma uygulaması geliştirilmesi, *Engineering Sciences (NWSAENS)*, 13(2), 119-127.
- Boz M., Canbazoglu E., Özen Z. ve Gülseçen S., (2018). Otel rezervasyon iptallerinin makine öğrenmesi yöntemleri ile tahmin edilmesi, *Veri Bilimi*, 1(1), 7-14.
- Breiman, L. (2001). Random forests. *Machine Learning*, 45, 5-32.
- Breiman, L., Friedman, J. H., Olshen, R. A. ve Stone, C. J. (1984) *Classification and Regression Trees*, Wadsworth, Belmont, CA. Republished by CRC Press.
- Caicedo-Torres W. ve Payares F. (2016). A Machine Learning Model for Occupancy Rates and Demand Forecasting in the Hospitality Industry. Ed. Montes y Gómez M., Escalante H., Segura A., Murillo J. In *Advances in Artificial Intelligence - IBERAMIA 2016*. IBERAMIA 2016. Lecture Notes in Computer Science, vol 10022. Springer, Cham.

- Chen, C. C., Schwartz, Z. ve Vargas, P. (2011). The search for the best deal: how hotel cancellation policies affect the search and booking decisions of deal-seeking customers. *International Journal of Hospitality Management*, 30, 129-135.
- Chen, C. C. ve Xie, K. (2013). Differentiation of cancellation policies in the U.S. hotel industry, *International Journal of Hospitality Management*, 34, 66-72.
- Çevik, K. K. ve Kayakuş, M. (2020). Bilişim teknolojileri departmanında kullanıcıların taleplerine cevap verme süresinin makine öğrenmesi ile tahmin edilmesi. *Mühendislik Bilimleri ve Tasarım Dergisi*, 8(3), 728-739.
- Dreyfus, G. (2005). *Neural networks: methodology and applications*. Germany: Springer Science & Business Media.
- Falk, M. ve Vieru, M. (2018). Modelling the cancellation behaviour of hotel guests, *International Journal of Contemporary Hospitality Management*, 30(10), 3100-3116.
- Giudici, P. (2003). *Applied Data Mining: Statistical Methods for Business and Industry*. West Sussex, England: John Wiley & Sons.
- Hajibaba, H., Boztuğ, Y. ve Dolnicar, S. (2016). Preventing tourists from canceling in times of crises, *Annals of Tourism Research*, 60, 48-62.
- Ho, T. K. (1998). The random subspace method for constructing decision forests, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(8), 832-844.
- Kantardzic, M. (2019). *Data Mining Concepts, Models, Methods, and Algorithms Third Edition*. New Jersey: John Wiley & Sons, Inc.
- Kim, Y., Chong, J. W., Chon, K. H. ve Kim, J. (2013). Wavelet-based AR-SVM for health monitoring of smart structures. *Smart Materials and Structures*, 22, 1-12.
- Lu, W., Wang, W., Leung, A. Y., Lo, S. M., Yuen, R. K., Xu, Z. and Fan, H. (2002). Air Pollutant Parameter Forecasting Using Support Vector Machines. In *Proceedings of the 2002 International Joint Conference on Neural Networks. IJCNN'02 (Cat. No. 02CH37290) (Vol. 1, pp. 630-635)*. IEEE.
- Masiero, L., Viglia, G. ve Nieto-Garcia, M. (2020). Strategic consumer behavior in online hotel booking. *Annals of Tourism Research*, 83, 1-13.
- Morales, D. R. ve Wang, J. (2010). Forecasting cancellation rates for services booking revenue management using data mining. *European Journal of Operational Research*, 202, 554-562.
- Nikseresht, A., ve Ziarati, K. (2015). Review on the newest revenue management demand forecasting methods. In *International Conference on Management, Economics and Industrial Engineering*, 1(1).
- Nisbet, R., Miner, G. ve Yale, K. (2018). *Handbook of Statistical Analysis and Data Mining Applications (Second Edition)*. London: Academic Press, Elsevier Inc.
- 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 (pp. 154-168). Berlin, Heidelberg: Springer-Verlag.
- Palmer, A., Jiménez, R. ve Gervilla, E. (2011). *Data mining: machine learning and statistical techniques*. Ed. Funatsu, K. ve Hasegawa, K. In *Knowledge-Oriented Applications in Data Mining*. (pp. 373-396), India: Intech Open Access Publisher.
- Poongodi, M., Sharma, A., Vijayakumar, V., Bhardwaj, V., Sharma, A. P., Iqbal, R. ve Kumar, R. (2020). Prediction of the price of Ethereum blockchain cryptocurrency in an industrial finance system. *Computers and Electrical Engineering*, 81, 1-12.
- Rokach, L. ve Maimon, O. (2015). *Data mining with decision trees theory and applications (2nd edition)*. Singapore: World Scientific Publishing Co. Pte. Ltd.
- Sánchez-Medina, A. J. ve C-Sánchez, E. (2020). Using machine learning and big data for efficient forecasting of hotel booking cancellations, *International Journal of Hospitality Management*, 89, 1-9.

- Sapankevych, N. I. ve Sankar, R. (2009). Time series prediction using support vector machines: a survey. *IEEE Computational Intelligence Magazine*, 4, 24-38.
- Saputro, P. H. ve Nanang, H. (2021). Exploratory data analysis & booking cancelation prediction on hotel booking demands datasets, *Journal of Applied Data Sciences*, 2(1), 40-56.
- Steinberg, D. (2009). *CART: Classification and Regression Trees*, Ed. Wu, X. ve Kumar, V., In *The Top Ten Algorithms in Data Mining* (1st ed.). (pp. 179-201), Boca Raton, USA: Chapman and Hall/CRC, Taylor & Francis Group.
- Steinberg, D. ve Colla, P. (1995). *CART: Tree-Structured Non-Parametric Data Analysis*. San Diego, CA: Salford Systems.
- Uzbaş, B. (2017). Sayısal dental modellerden otomatik cinsiyet tespiti, (Yayımlanmamış Doktora Tezi), Selçuk Üniversitesi Fen Bilimleri Enstitüsü, Konya.
- Yakut, E., Elmas, B. ve Yavuz, S. (2014). Yapay sinir ağları ve destek vektör makineleri yöntemleriyle borsa endeksi tahmini. *Süleyman Demirel Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 19(1), 139-157.
- Wang, C., Shu, Q., Wang, X., Guo, B., Liu, P. ve Li, Q. (2019). A random forest classifier based on pixel comparison features for urban LiDAR data. *ISPRS Journal of Photogrammetry and Remote Sensing*, 148, 75-86.