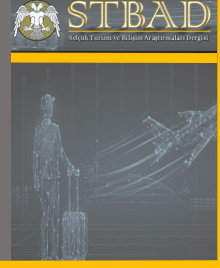




STBAD

Selçuk Turizm ve Bilişim Araştırmaları Dergisi
22, 1(2): 15-24



Makale Bilgisi

Gönderilme Tarihi: 20.08.2022

Kabul Tarihi: 29.08.2022

Araştırma Makalesi

Otel Rezervasyon İptal Tahmin Modelinin Veri Madenciliği Algoritmaları ile Uygulanması

Dr. Öğr. Üyesi Kevser Şahinbaş, İstanbul Medipol Üniversitesi, İşletme ve Yönetim Bilimleri Fakültesi, İstanbul, E-posta: ksahinbas@medipol.edu.tr
ORCID: <https://orcid.org/0000-0002-8076-3678>

Dr. Öğr. Üyesi Özge Doğuç, İstanbul Medipol Üniversitesi, İşletme ve Yönetim Bilimleri Fakültesi, İstanbul, E-posta: odoguc@medipol.edu.tr
ORCID: <https://orcid.org/0000-0002-5971-9218>

Öz

Otel rezervasyon iptalleri, oda rezervasyon sistemleri üzerindeki etkileri nedeniyle gelir yönetiminde kritik olarak kabul edildiği için otel ve konaklama sektörünün önemli bir problemidir. Gelir yöneticilerinin bakış açısından, satılan her otel odası, uçak koltuğu vb. ek kar sağlamaktadır. Bu nedenle, satılmayan her oda ya da koltuğun potansiyel bir kaybı temsil ettiği sonucu çıkmaktadır. Müşterileri iptal etmeye sevk eden sebepler veya bunun nasıl önlenebileceği hakkında bilgi sahibi olmak oldukça önem taşımaktadır. Bu çalışmanın amacı bireysel otel iptallerinin tahminine odaklanmak ve iptal üzerinde en çok hangi parametrelerin olduğunu ortaya çıkarmaktır. Bu çalışmada Veri Madenciliği tekniklerinden Karar Ağaçları ve Rastgele Orman algoritmaları uygulanmıştır. Elde edilen sonuçlara göre %86.7 oranında doğruluk oranı ile Rastgele Orman algoritması daha iyi sonuç vermiştir. Depozito tipi ve müşterinin daha önce rezervasyon iptali yapıp yapmadığı parametrelerinin sınıflandırma üzerinde en fazla etkiye sahip oldukları gözlemlenmiştir. Bu modeli benimseyen kuruluşlar turist varışlarıyla ilgili eylem protokollerini güçlendirebilir, rezervasyon yönetim sistemleri ve iptal politikalarını optimize edilebilirler.

Anahtar Kelimeler: Otel Rezervasyonları, İptaller, Gelir Yönetimi, Veri Yönetimi, Tahminleme

Selçuk Turizm ve Bilişim Araştırmaları Dergisi ©

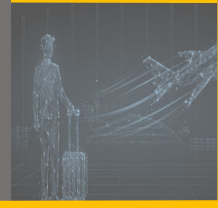
Önerilen Atıf:

Şahinbaş, K. ve Doğuç, Ö. (2022). Otel rezervasyon iptal tahmin modelinin veri madenciliği algoritmaları ile uygulanması. *Selçuk Turizm ve Bilişim Araştırmaları Dergisi*, 1(2):15-24.



STBAD

*Selcuk Tourism and Information Technology Research
Journal, 22, 1(2): 15-24*



Article Information
Received 20.08.2022
Accepted 29.08.2022
Research Article

Hotel Reservation Cancellation Prediction Model with Data Mining Algorithms

Kevser Şahinbaş, İstanbul Medipol University [e-mail: ksahinbas@medipol.edu.tr](mailto:ksahinbas@medipol.edu.tr)
ORCID: <https://orcid.org/0000-0002-8076-3678>

Özge Doğuç, İstanbul Medipol University, [e-mail: odoguc@medipol.edu.tr](mailto:odoguc@medipol.edu.tr)
ORCID: <https://orcid.org/0000-0002-5971-9218>

Abstract

Hotel reservation cancellations are an important aspect of the hotel and hospitality industry as they are considered critical in revenue management due to their impact on room reservation systems. From the point of view of revenue managers, every unit sold includes hotel rooms, airplane seats, etc. provides additional profit. Therefore, it follows that each unsold unit represents a potential loss. It is very important to be informed about the reasons that lead customers to cancel or how this can be avoided. The aim of this study is to focus on the prediction of individual hotel cancellations and to reveal which parameters have the most on cancellation. In this study, Decision Trees and Random Forest algorithms from Data Mining techniques were applied. According to the results obtained, the Random Forest algorithm achieved better results with an accuracy rate of 86.7%. It has been observed that out of the 28 parameters of a reservation, the deposit type and whether the customer has canceled a reservation before have the most influence on the classification. Organizations that adopt this model can strengthen action protocols on tourist arrivals, optimize reservation management systems and cancellation policies.

Keywords: Hotel Reservations, Cancellations, Revenue Management, Data Management, Forecasting

Selcuk Tourism and Information Technology Research Journal ®

Suggested Citation:

Sahinbas, K. ve Doguc, Ö. (2022). Otel rezervasyon iptal tahmin modelinin veri madenciliği algoritmaları ile Uygulanması. *Selcuk Tourism and Information Technology Research Journal*, 1(2):15-24.

1. GİRİŞ

Turizm dünyadaki en yaygın endüstrilerden biridir ve küresel ekonomideki önemi yadsınamaz bir gerçektir. 2019'da uluslararası turist sayısının Amerika'da 216.3, Asya Pasifik'te 360.1, Orta Doğu'da 73, Afrika'da 68.1 ve Avrupa'da 745 milyona yükseldiği düşünüldüğünde büyümenin bir kanıtı olarak görülebilmektedir (Statista, 2022). Bu büyümenin bir parçası olmak turizm ve seyahatle ilgili endüstrilerde, gelir yönetimi talep tahmini ve tahmin sorunları üzerine doğru kararlar almayı gerektirmektedir. Konaklama endüstrisinde gelir yönetimi, doğru odayı doğru misafir için doğru fiyatla, doğru zamanda ve doğru kanaldan erişilebilir kılmak olarak tanımlanmaktadır (Mehrotra ve Ruttley, 2006). Otel yöneticilerinin karşılaştığı en önemli sorunlardan biri, otelin kapasitesini taleple eşleştirmeye çalışmaktır. Konaklama endüstrisinde ürün stoklanamamakta olup bu da otelcileri belirli bir zaman diliminde talep ve sınırlı sayıda oda ile başa çıkmaya zorlamakta, böylece boş her oda gelir kaybına neden olmamaktadır. Bu durum otel yöneticilerinin gelecekteki taleple ilgilenmeyle ilgili olan doluluk oranını en üst düzeye çıkararak gelirlerini artırmaya çalışmasına yol açmaktadır.

Otel gelir yönetimini en üst düzeye çıkarmanın ana yönlerinden biri, sektörün maruz kaldığı belirsiz ortam nedeniyle kolay bir iş olmamakla birlikte, mevcut odalar için organizasyon ve planlama prosedürlerinin verimliliğinde yatmaktadır. Ekonomik kriz, enflasyon, çevresel değişiklikler, savaşlar, düzenleyici değişiklikler, yeni müşteri talebi veya teknolojik değişiklikler yönetimin bu sektörde dikkate alınması gereken faktörlerdir (Yüksel, 2005).

Rezervasyon sistemleri müşterilere hizmet alımını garanti altına alma imkânı, otellere de kaynaklarını doğru yönetme olanağı sunar. Ancak hizmet zamanına yakın yapılan iptaller, özellikle yüksek bir gelir kaybına neden olabileceğinden, bir iptal için bildirim süresi kritik bir husus olarak düşünülmelidir (Chen vd., 2011). İptaller, geliri doğrudan etkilediği ve dolayısıyla gelir yönetiminde hayati önem olarak kabul edildiği için otel ve konaklama sektörü üzerinde önemli bir etkiye sahiptir. Bazı araştırmalara göre iptaller gelirin yaklaşık %20'sini oluşturan bir gelir kaybını oluşturabilmektedir (Sierag vd., 2015). Bu nedenle, organizasyon ve planlama esas olmakla birlikte doğru bir tahmin aracına sahip olmak çok önemlidir (Pereira, 2016). Rezervasyon iptallerinden kaynaklanan gelir kaybının ortadan kaldırılması için iptallerin önceden tahmin edilebilmesi hayati önem taşımaktadır.

Turizm ve seyahatle ilgili endüstrilerde, gelir yönetimi talep tahmini ve tahmin sorunları üzerine yapılan araştırmaların çoğu, havacılık endüstrisinden, Yolcu Adı Kaydı (PNR) olarak bilinen formatta verileri kullanmaktadır. Bu durum havacılık endüstrisi tarafından geliştirilmiş bir formattır. Ancak, konaklama, seyir, tema parkları vb. gibi turizm ve seyahat endüstrileri, endüstrinin özel verileri olmadan tam olarak keşfedilemeyecek farklı gereksinimlere ve özelliklere sahiptir. Bu nedenle, bu sınırlamanın üstesinden gelmeye yardımcı olmak için talep verileriyle iki otel veri kümesi paylaşılmıştır. Bu çalışmada bir otel verisi kullanılarak iptal tahmini yapılmıştır.

Turizm ve otel talep tahmin ve iptalleri ile ilgili bazı çalışmalar sunulmaktadır. Sánchez vd (2020) tarafından yapılan çalışmada PNR verileri aracılığıyla Yapay Zekâ tekniklerini kullanılarak kısa bir zaman diliminde iptal etme olasılığı bulunan kişileri belirlemek amaçlanmıştır. 7 gün öncesinden yapılan iptallerde %80 doğruluk değerini elde etmişlerdir. Sánchez ve Eleazar (2020) çalışmalarında 13 bağımsız değişken kullanarak otel rezervasyon iptallerinin tahminini yapan model önermektedir. Çalışmalarında genetik algoritmalarla optimize edilmiş yapay sinir ağlarının yanı sıra makine öğrenimi teknikleri uygulanarak %98'e varan bir iptal oranı elde etmişlerdir. Wu vd. (2017) turizm ve otel talep tahmininde genetik algoritmaların, Gaussian Process Regresyonun ve bulanık yöntemlerin bu alanda uygulanmış olmalarına rağmen, Yapay Sinir Ağlarının (YSA) en sık kullanılan yöntem olduğunu belirtmişlerdir. Otel rezervasyonlarının iptalinin tahmin edildiği çalışmada Antonio vd. (2017) Güçlendirilmiş Karar Ağacı, Rastgele Ormanı, Yerel Olarak Derin Destek Vektör Makinesi ve Sinir Ağı; Algarve (Portekiz) tatil bölgesinde bulunan dört otelden alınan veri setlerini kullanarak iptal oranlarını tahmin etmek amacıyla dört farklı otel veri seti kullanmışlardır. Çalışmada her otel için ayrı modeller oluşturulmuş ve ortalama olarak %90'ın üzerinde bir doğruluk değeri bulunmuştur. Romero Morales ve Wang (2010), veri madenciliği yöntemlerinden karar ağacı tabanlı yöntemler, Naive Bayes tabanlı yöntemler ve destek vektör makinesini kullanarak gelir yönetimi rezervasyonu yapan hizmetler için iptal oranlarını tahmin etmişlerdir. Rezervasyonun fiyatı, oda tipi, rezervasyonun yapıldığı kanal veya kullanılan rezervasyon sistemi gibi 14 değişken kullanmışlardır. Otel iptali tahmini için ağaç tabanlı yöntemler ve çekirdek yöntemlerinin en popüler olduğunu belirtmişler, özellikle Destek Vektör Makinesi'nin bu amaçla kullanılan en dikkate değer yöntem olduğunu belirtmişlerdir. Ayrıca rezervasyon hizmet zamanına ne kadar yakınsa, test edilen her teknikte o kadar az hata bulunması sonucuna da varmışlardır.

Bu çalışmada sınıflandırma algoritmalarından C4.5 Karar Ağacı ve Rastgele Orman Algoritması kullanılarak modeller geliştirilmiştir. Rastgele Orman metodunun doğruluk oranı %86,7 olarak bulunmuştur. Rezervasyon Durumu değişkenini en çok etkileyen 2 nitelik olarak depo tipi ve müşterinin daha önce rezervasyon iptali yapıp yapmadığı sonucu elde edilmiştir.

2. YÖNTEM

Bu çalışma, otel rezervasyonlarından gelen verileri kullanmakta ve sınıflandırma algoritmaları yardımıyla rezervasyon iptalleri ve no-show durumlarını doğru bir şekilde tahmin edebilen bir veri madenciliği modeli oluşturmaktadır. Sınıflandırma, bir otel rezervasyonunu Normal veya İptal Edildi/No-Show olarak kategorize etme yöntemidir. Bu çalışma, müşterilerin bir rezervasyonu rastgele iptal etmediğini varsayar ve bu şekilde bir rezervasyonun iptal/no-show mu yoksa rezervasyonun kullanılmasıyla mı sonuçlanacağını etkili ve doğru bir şekilde tahmin etmek mümkündür. Çalışmada rezervasyon verilerini analiz etmek ve rezervasyon durumlarını doğru bir şekilde tahmin edebilen veri modelleri oluşturmak için literatürde mevcut olan sınıflandırma ve tahmin algoritmaları kullanılmıştır.

2.1. Karar Ağacı Algoritması

Karar ağaçları, konuşma tanımadan tıbbi tanıya kadar birçok farklı alanda kullanılan bir algoritmadır (Safavian ve Landgrebe, 1991). Birçok alanda tercih edilmesinin en büyük nedenlerinden biri, sınıflandırıcıların karmaşık karar verme kuralları koymaması, bu nedenle hızlı bir model elde edilmesi ve modelin çıktısının diğer algoritmalara göre daha kolay yorumlanmasıdır. Bir karar ağacı algoritmasını oluşturan unsurlardan biri olan kök düğüm, sınıfların etiketlerini içerirken, karar ağacındaki yapraklar köke bağlı olmayanlar olarak adlandırılır (Safavian ve Landgrebe, 1991). Karar ağaçlarının çalışması sırasıyla köke inerek verileri bölerek öğrenmeyi sağlamak ve birden fazla algoritma ile öğrenme gerçekleştirebilen bir yapıya sahip olmaktır (Brijain vd., 2014). Geliştirilen algoritmalar arasında öznitelikler arasındaki iletişimi sınıflandıran CHID algoritması ve Gini ile karar ağacı oluşturduktan sonra en doğru özniteliği seçen CART algoritması gibi veri kümesine uygun birçok algoritma kullanılmaktadır. Çalışmada kullanılan karar ağacı uygulaması, entropi çalıştıran ID3 algoritmasıdır. Bu algoritmanın seçilmesinin nedenlerinden biri çalışmada kullanılan veri setinde çok fazla kategorik veri bulunmasıdır. Genel olarak kategorik değişkenlerde kullanılmak üzere entropi ve Gini tabanlı algoritmalar kullanılır ve entropi tabanlı algoritmalarından biri olduğu için ID3 tercih edilir. ID3 ağaç yapısındaki düğümler için en uygun öznitelige karar verebilmek için bilgi kazancı entropi ile desteklenir ve en düşük entropi belirsizliğine sahip öznitelik en uygun olarak seçilir (Brijain vd., 2014). Entropi, özniteliklerin oluşturduğu farklılıkları karar ağacından yukarıdan aşağıya toplar ve destekleyici özniteliklerle çıkarımlar yapar, burada amaç, karar ağacı bölme işlemini gerçekleştirirken en düşük entropi düzeyini bulmaktır (Li ve Claramunt, 2006). Bilgi kazancı, entropideki azalma oranını ölçer ve dolaylı olarak en düşük entropi düzeyine sahip özellik en büyük bilgi kazancına sahiptir ve bu özellik karar ağacında düğüm olarak seçilir. Diğer bir karar ağacı algoritması C4'tür ve bu algoritma kategorik ve sayısal verileri işleyebilen ID3 uzantılı bir sınıflandırıcı algoritmadır (Brijain vd., 2014). Karar ağacı algoritmasında kullanılabilecek çoklu yöntemler arasından veri tipine, büyüklüğüne ve veri setinin yapısına göre en uygun olanı bulmak ve seçmek en doğru tahmin sonucunu sağlar.

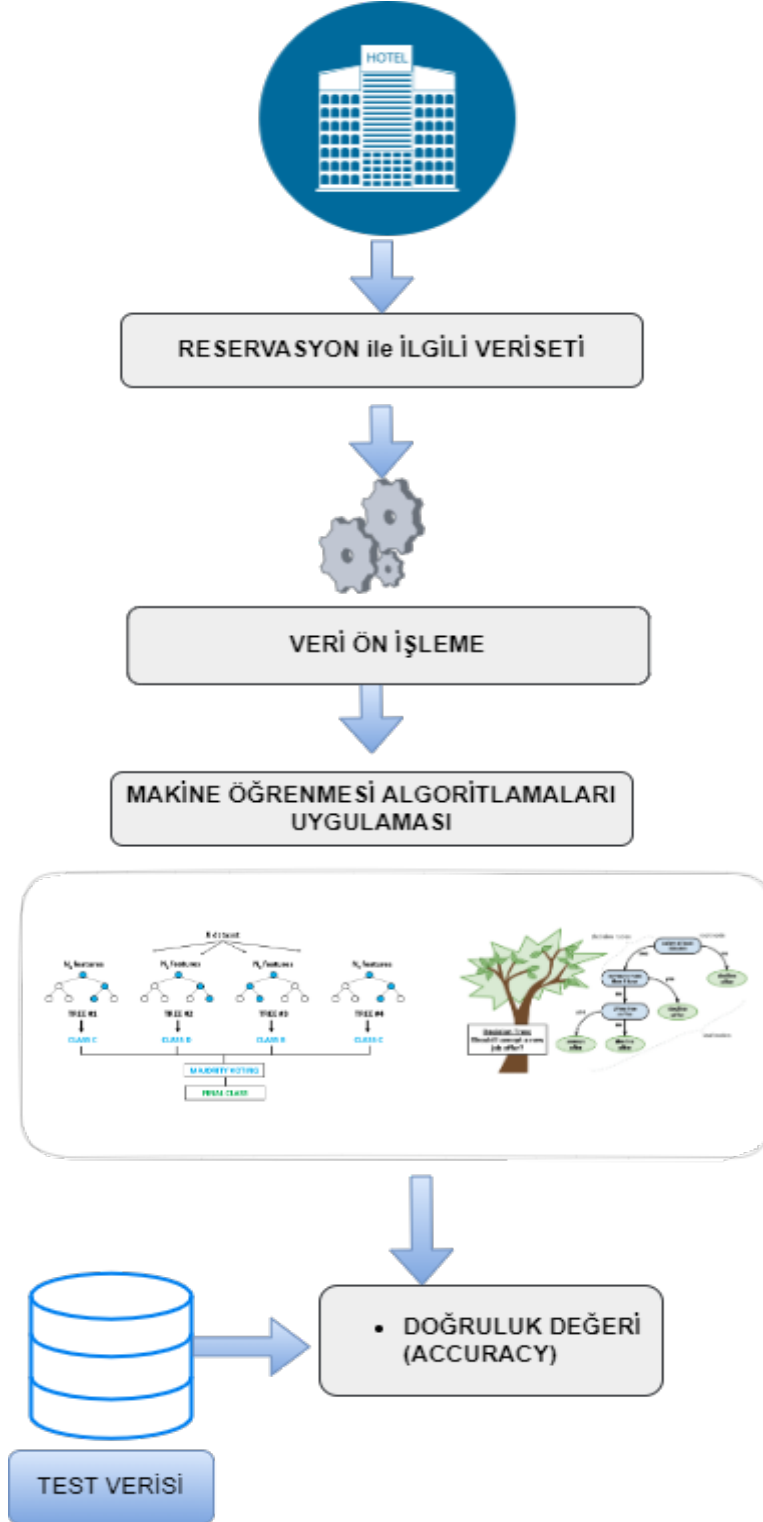
2.2. Rastgele Orman Algoritması

Rastgele Orman (Random Forest- RF) algoritması Breiman (2001) tarafından yapılandırılmış bir karar ağacı koleksiyonundan oluşan bir sınıflandırıcı olarak tanımlanmıştır. Rastgele orman algoritmasının ortaya çıkmasının nedenlerinden biri, karar ağaçlarının aşırı uyum sorunu yani test setindeki başarısız performansdır ve bu durumda dolaylı olarak genel tahmini olumsuz yönde etkiler (Schonlau ve Zou, 2020). Genel doğruluğu artırmak için birden fazla ağaç oluşturulabilir ve altuzay kümeleri dikkate alınarak oran artırılabilir (Schonlau ve Zou, 2020). Bunun üzerine birden fazla önyükleme örneğinden oluşan ağaçlardan alınan tahminlerin ortalamasını oluşturan rastgele orman algoritması ortaya çıkar. Karar ağacı ve algoritmada ne kadar çok karar ağacı olursa, tahmin başarısı o kadar iyi olur. Özellik seçimini otomatik olarak yapar ve bunun için ayrı bir algoritma kullanmasına gerek kalmaz, bunun sonucunda bir tahmin üretir (Lin vd., 2017). Rastgele Orman hem regresyon hem de sınıflandırma problemleri için kullanılabilen bir algoritmadır (Liaw ve Wiener, 2002). Veri setinden önce n örnekten oluşan bir bootstrap seti oluşturulur, ardından örnekler tek tek ele alınarak bir sınıflandırma ağacı oluşturulur ve düğümlerde tahmin edicilerin rastgele örnekleri yeniden örneklenir. Bu olaydan en iyi hangisi ayrılırsa o seçilir (Liaw ve Wiener, 2002). Modelde kullanılacak ağaçlar oluşturulurken CART algoritması kullanılmakta ve diğer sınıflandırıcılara göre veri setindeki gürültüyü yönetme ve sınıflandırma başarısı ile iyi bir performans gösterdiği bilinmektedir (Liu, 2014). Bu algoritmanın torbalama yöntemi, genel hatayı tahmin etme, sınıflandırmanın eşit dağılımını sağlama, büyük veride gürültünün iyi yönetimi ve analistler için kolay yorumlama gibi özellikleri ile büyük avantajlar

sağlamaktadır. Bir dezavantaj olarak sınıflandırıcılar bir süre sonra güncelliğini yitirecek ve verileri güncellemede çok hızlı oldukları için tahmin doğruluğu azalacaktır (Liu, 2014).

2.3. Tahmin Model Şeması

Otel rezervasyon iptal tahmini için veriler açık erişilebilir web sitesinden elde edilmiştir. Genel mimarinin şeması Şekil 1'de gösterilmiştir. İlk olarak, veriler ön işlemden geçmektedir. Bu aşamada veriler kodlanmıştır. Daha sonra analiz yapmak için veriler bir csv dosyasına gönderilir. Veriler normalize edildikten sonra verilere Rastgele Orman ve Karar Ağaçları algoritmaları uygulanmıştır. Sonuç olarak, model tahmin edilen çıktıları üretmiştir.



Şekil 1. Otel Rezervasyon İptal Tahmin Model Şeması

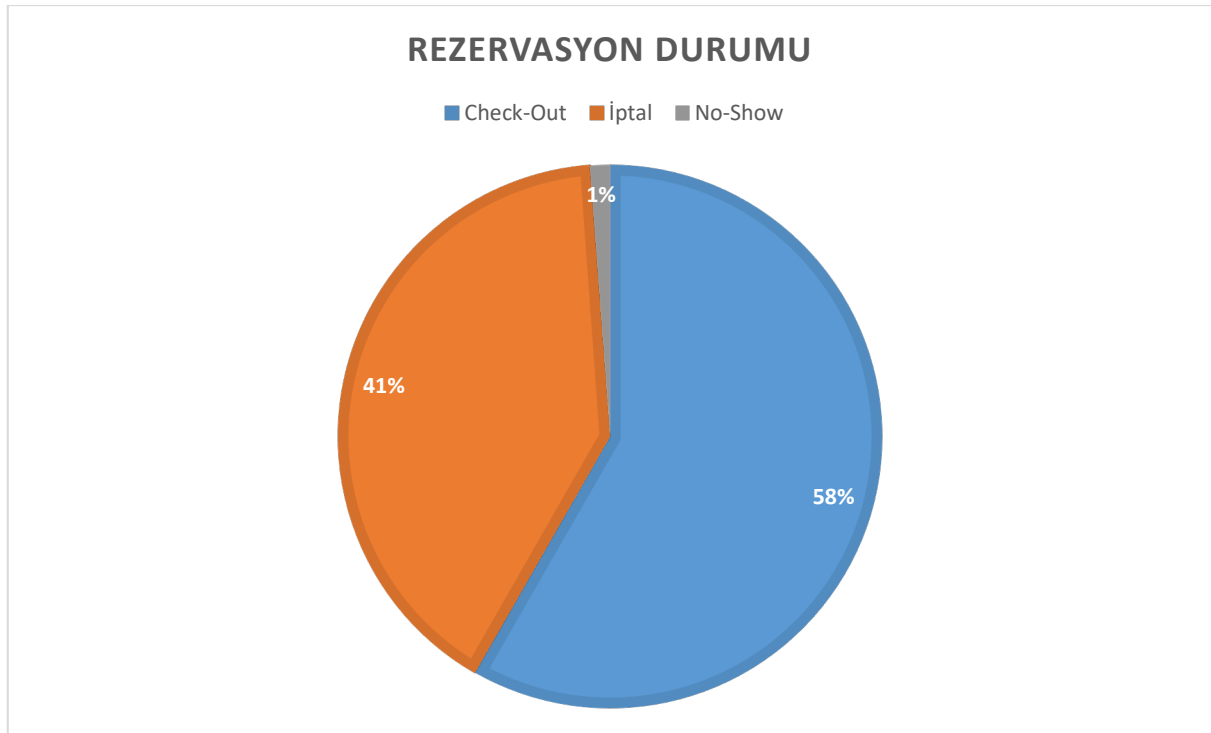
2.4. Veri Kümesi

Bu çalışmada kullanılan veri kümesi, bir otel rezervasyonu ile ilgili 28 farklı niteliği içermektedir. Bir otelin rezervasyon bilgilerini içeren veri kümesi halka açık bir platformdan elde edilmiştir. Nitelik listesi Tablo 1'de verilmiştir.

Tablo 1. Veri Kümesinde Yer Alan Niteliklerin Listesi

LeadTime (Erken Rezervasyon Süresi)	Meal (Yemek Durumu)
ArrivalDateYear (Varış Yılı)	Country (Ülke)
ArrivalDateMonth (Varış Ayı)	MarketSegment (Pazar Segmenti – Online, telefon vs.)
ArrivalDateWeekNumber (Varış Haftası)	DistributionChannel (Dağıtım Kanalı)
ArrivalDateDayOfMonth (Varış Günü)	IsRepeatedGuest (Tekrarlı Müşteri Mi?)
StaysInWeekendNights (Kalınan Hafta Sonu Günü)	PreviousCancellations (Önceki İptal Sayısı)
StaysInWeekNights (Kalınan Hafta İçi Günü)	PreviousBookingsNotCanceled (Önceki İptal Edilmeyen Rezervasyon Sayısı)
Adults (Yetişkin Sayısı)	ReservedRoomType (Oda Tipi)
Children (Çocuk Sayısı)	AssignedRoomType (Verilen Oda Tipi)
Babies (Bebek Sayısı)	BookingChanges (Rezervasyon Değişiklikleri)
CustomerType (Müşteri Tipi – Grup, Acenta vs)	DepositType (Depozito Türü)
RequiredCarParkingSpaces (Araç Sayısı)	Agent (Rezervasyonu Alan Kişi)
TotalOfSpecialRequests (Özel İstek Sayısı)	Company (Rezervasyonu Alan Şirket)
ReservationStatus (Rezervasyon Durumu)	DaysInWaitingList (Beklemede Geçen Gün Sayısı)

Bu nitelikler arasında *ReservationStatus* hedef değişkendir. *ReservationStatus* niteliği şu üç değerden birini alabilir: Check-Out, İptal veya No-Show. Çalışmada kullanılan veri kümesi yaklaşık 80 bin rezervasyon içermektedir ve bunların %41'i *İptal* ile sonuçlanmıştır. Şekil 2, veri kümesindeki rezervasyon durumlarının dağılımını göstermektedir.



Şekil 2. Veri setinin Rezervasyon Durumu'na göre dağılımı

2.5. Model Seçimi

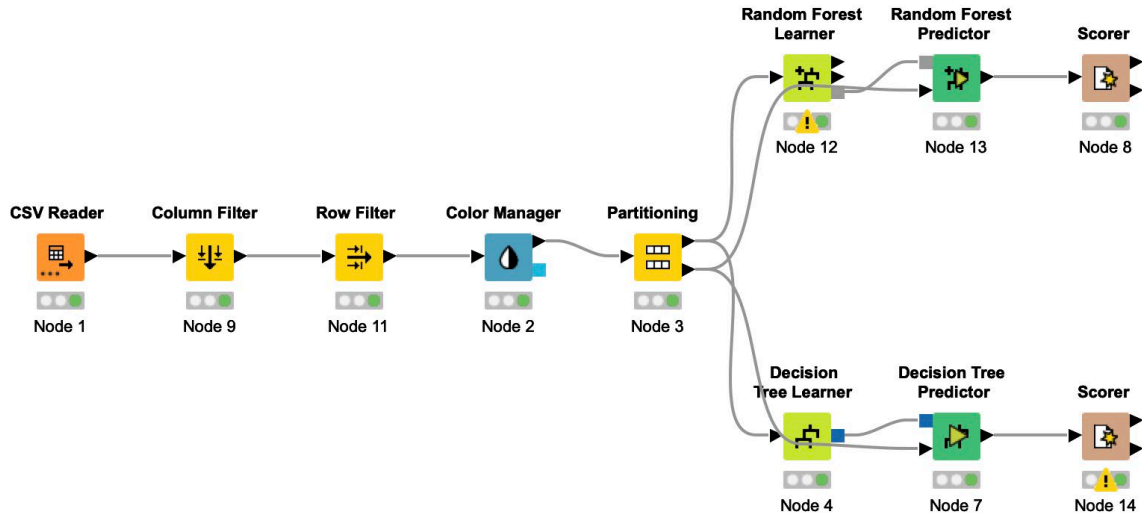
Makine öğrenmesinde, sınıflandırma yöntemlerinde kurulan modelin performansını ölçen çeşitli ölçüler bulunmaktadır. Bu çalışmada, model performansı için genel doğruluk kullanılmıştır (Han ve Kamber, 2006).

3. BULGULAR

Bu araştırma, 80.000 civarı rezervasyon verisinin rezervasyon kaydı içeren bir otel tarafından sağlanan verilere farklı makine öğrenimi teknikleri uygulanarak gerçekleştirilmiştir.

Bu çalışmada çeşitli veri madenciliği metotları kullanılmış ve rezervasyon durumunu doğru tahmin etme sonuçları karşılaştırılmıştır. Bu amaçla, Knime platformunda bir iş akışı oluşturulmuştur. Knime, çok sayıda veri analizi ve veri madenciliği paketini içeren ücretsiz bir platformdur. Farklı tiplerde veri kaynaklarından gelen verinin ayıklanması, dönüştürülmesi ve yüklenmesi (ETL) sağladığı gibi, Python, Java gibi programlama dillerinde yazılmış veri madenciliği ve makine öğrenmesi algoritmalarıyla veri işlemeye de yardımcı olur.

Bu çalışma için, Knime platformu üzerinde 2 farklı veri madenciliği metodu denenmiştir: Decision Tree (Karar Ağacı) ve Random Forest (Rastgele Orman). İki metot için de bir learner (öğrenici) ve bir de predictor (tahmin edici) paket kullanılmış ve sonuçlar karşılaştırılmıştır. Knime platformu üzerinde oluşturulan iş akışı Şekil 3'te verilmiştir.



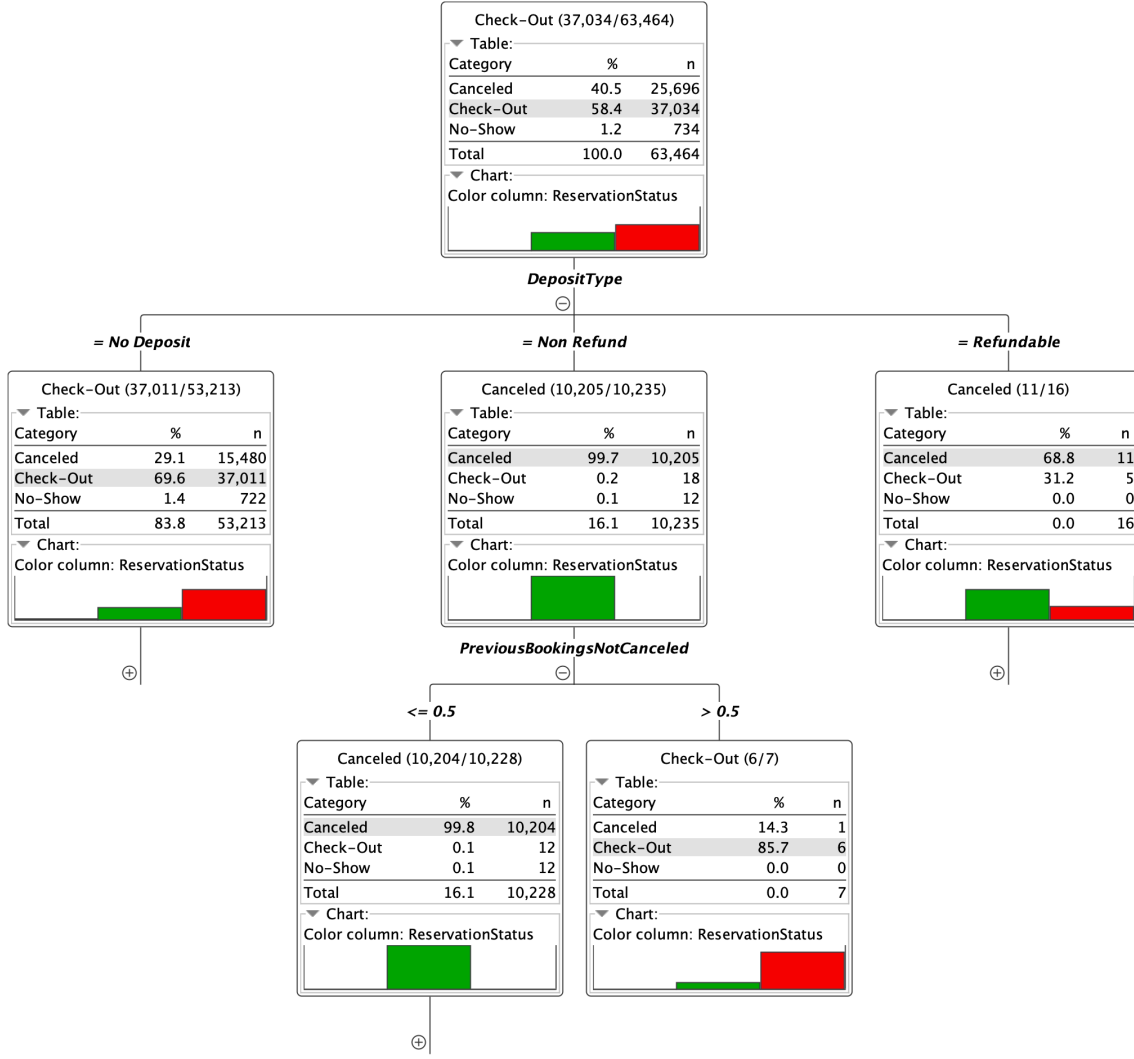
Şekil 3. Knime platformunda oluşturulan iş akışı

İş akışının ilk adımında CSV formatında gelen veri Knime'a yüklenir ve sonrasında boş (null) veri içeren satır ve sütunlar silinir. Sonraki adımda, veri training (eğitim) ve validation (doğrulama) amaçlı iki kümeye ayrılır. Eğitim kümesi veri madenciliği metotlarının verinin yapısını öğrenmesinde kullanılır ve bu çalışmada toplam verinin %80'i eğitim kümesine ayrılmıştır. Kalan %20'lik veri ise doğrulama amaçlı kullanılmıştır. Doğrulama sonuçları bir sonraki kısımda detaylandırılmıştır. Eğitim ve doğrulama kümelerinin ayrıştırılması rastgele yapılmıştır.

3.1. Sonuçların Karşılaştırılması

Önceki bölümde de bahsedildiği gibi, bu çalışmada Karar Ağacı ve Rastgele Orman yöntemleri kullanılmış ve oluşan tahminlerin doğrulukları karşılaştırılmıştır.

Şekil 4'te eğitim (training) kümesiyle Karar Ağacı metodunun oluşturduğu ağaç yapısı verilmiştir. Buna göre, Rezervasyon Durumu değişkenini en çok etkileyen 2 nitelik Depozito Tipi ve müşterinin daha önce rezervasyon iptali yapıp yapmadığıdır. Örneğin, geri ödemeli rezervasyon yapan bir müşteri daha önce de rezervasyon iptali yaptıysa, bu rezervasyonu da iptal etme olasılığı %99,8'e çıkmaktadır.



Şekil 4. Karar ağacı metodunun oluşturduğu ağaç yapısı

Bu ağaç yapısı kullanılarak Karar Ağacı metodunun doğrulama verisini kullanarak %83,7 doğrulukla rezervasyon durumunu tahmin ettiği görülmektedir. Karar Ağacı metodu için ortaya çıkan karmaşıklik matrisi (confusion matrix) Şekil 5'te verilmiştir.

Row ID	Check-Out	Canceled	No-Show
Check-Out	7932	1194	26
Canceled	1201	5258	19
No-Show	113	32	37

Şekil 5. Karar ağacı metodu için oluşan karmaşıklik matrisi

Diğer taraftan Rastgele Orman metodu, 100 farklı karar ağacı oluşturup içlerinden en tutarlı sonuçları veren ağaç yapısını kullanmaktadır. Dolayısıyla Karar Ağacı metoduna göre daha yüksek bir tahmin oranı sağlaması beklenebilir. Nitekim, Rastgele Orman metodunun doğruluk oranı %86,7 olarak bulunmuştur. Analiz sonucunda elde edilen bulgulara göre, Rasgele Orman modeli rezervasyon iptal tahmininde bariz bir avantaja sahiptir ve %86.7 doğruluk oranı ile otel rezervasyon tahmininde en yüksek sınıflandırma performansını elde etmektedir. Analiz sonucunda ortaya çıkan karmaşıklik matrisi Şekil 6'da verilmiştir.

Row ID	Check-Out	Canceled	No-Show
Check-Out	8587	603	4
Canceled	1358	5132	0
No-Show	122	24	36

Şekil 6- Rastgele Orman metodu için oluşan karmaşıklık matrisi

4. TARTIŞMA, SONUÇ ve ÖNERİLER

Rezervasyon iptalleri, otel gelirini doğrudan etkilediği ve bu nedenle gelir yönetiminde çok önemli yere sahip olduğu için otel ve konaklama sektörü üzerinde önemli bir etkiye sahiptir. No-show ve rezervasyon iptalleri oteller için önemli maddi kayba yol açmaktadır. Bu durumla başa çıkabilmek için oteller overbooking (çifte rezervasyon) yapmakta, ancak yoğun sezonlarda çakışan rezervasyonlardan kaynaklı müşteri memnuniyetini düşüren sorunlar çıkmaktadır. Oteller, yapılan rezervasyonların iptal veya no-show'la sonuçlanabilme ihtimaline karşılık genel önlemler alabilmekte; fakat rezervasyon özelinde bir öngörü yapmakta zorlanmaktadırlar.

Bu çalışmada, 80,000 civarı rezervasyon verisinin analizi neticesinde, rezervasyon durumunu %86,7 doğruluk oranıyla tahmin edebilen bir veri madenciliği modeli geliştirilmiştir. Bunun için farklı veri madenciliği metotları kullanılmış ve içlerinden en yüksek doğrulukla tahmin yapan Rastgele Orman metodu seçilmiştir.

İptallerin otel zincirleri üzerindeki etkisi, özellikle iptalleri azaltmak amacıyla tasarlanmış stratejilerin geliştirilmesi gerektiği sonucuna varılmıştır. Bu araştırma, otel rezervasyon kayıtları kullanarak otel iptallerini tahmin etmek için makine öğrenimine dayalı bir model önererek otel ve konaklama endüstrisine katkıda bulunmaktadır. Bu tahmin modelini bünyelerinde uygulamak isteyen kuruluşlar rezervasyon yönetim sistemleri ve iptal politikalarını güncelleyebilirler.

Gelecekteki çalışmalar için, diğer veri madenciliği algoritmaları ile tahmin yapılması, çalışmaya öznelik mühendisliği (feature engineering) ve öznelik önemi (feature importance) eklenmesi planlanmaktadır.

KAYNAKÇA

- Antonio N, de Almeida A, Nunes L. (2017). Predicting hotel booking cancellations to decrease uncertainty and increase revenue, *Tour. Manag. Stud.*, 13(2).
- Breiman L. (2001). Random forests. *Mach. Learn.*, 45(1), 5– 32.
- Brijain, M., Patel, R., Kushik, M. R., & Rana, K. (2014). A survey on decision tree algorithm for classification. *IJEDR* .
- Chen, C. C., Schwartz, Z., & 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(1), 129-135.
- Han, J., ve Kamber, M. (2006). *Data mining: concepts and techniques*. USA: Morgan Kaufmann. <https://www.statista.com/statistics/186743/international-tourist-arrivals-worldwide-by-region-since-2010/> Erişim Tarihi: 10.05.2022
- Liu, Y. (2014). Random forest algorithm in big data environment. *Computer modelling & new technologies*, 18(12A), 147-151.
- Li, X., & Claramunt, C. (2006). A spatial entropy-based decision tree for classification of geographical information. *Transactions in GIS*, 10(3), 451-467.
- Liaw, A., & Wiener, M. (2002). Classification and regression by randomForest. *R news*, 2(3), 18-22.
- Lin, W., Wu, Z., Lin, L., Wen, A., & Li, J. (2017). An ensemble random forest algorithm for insurance big data analysis. *IEEE Access*, 5, 16568-16575.
- Mehrotra R, Ruttley J. (2006). *Revenue management*. (2nd ed.). Washington, DC, American Hotel and Lodging Association.
- Pereira, L. N. (2016). An introduction to helpful forecasting methods for hotel revenue management. *International Journal of Hospitality Management*, 58, 13-23.

- Sánchez-Medina, A. J., & Eleazar, C. (2020). Using machine learning and big data for efficient forecasting of hotel booking cancellations. *International Journal of Hospitality Management*, 89, 102546.
- Sierag, D. D., Koole, G. M., van der Mei, R. D., Van der Rest, J. I., & Zwart, B. (2015). Revenue management under customer choice behaviour with cancellations and overbooking. *European journal of operational research*, 246(1), 170-185.
- Safavian, S. R., & Landgrebe, D. (1991). A survey of decision tree classifier methodology. *IEEE transactions on systems, man, and cybernetics*, 21(3), 660-674.
- Schonlau, M., & Zou, R. Y. (2020). The random forest algorithm for statistical learning. *The Stata Journal*, 20(1), 3-29.
- Sánchez, E. C., Sánchez-Medina, A. J., & Pellejero, M. (2020). Identifying critical hotel cancellations using artificial intelligence. *Tourism Management Perspectives*, 35, 100718.
- Yüksel, S. (2005). An integrated forecasting approach for hotels. In *International Symposium on the Analytic Hierarchy Process (ISAHP)* (Vol. 10).
- Wu, D. C., Song, H., & Shen, S. (2017). New developments in tourism and hotel demand modeling and forecasting. *International Journal of Contemporary Hospitality Management*.