

Proje kapsamında;

Problem nedir?

Konaklama alanında, doğru odaya, doğru misafiri doğru zamanda ve doğru kanaldan erişilebilir duruma getirmek gelir yönetimini önemli bir ölçüde etkilemektedir.

Proje kapsamında karşılaşılan problem otelin kapasitesini taleple eşleştirmeye çalışmaktır. Bu zor durumla başa çıkmaya çalışan otel yöneticileri başka bir problem olarak da boş her odanın gelir kaybına neden olduğu gözlemlenmiştir.

Örneğin çevresel faktörlerin etkileri (enflasyon, savaşlar, ekonomik kriz) gibi durumlar sektörü olumsuz etkileyen durumlar arasında yer alır.

Bildiğimiz üzere oteller rezervasyon sistemlerini kullanır, müşterilerin hizmet alımını garanti altına alma imkanı da otellerin kaynaklarının doğru yönetme olanağı sunar. Ancak hizmet alımına yakın bir sürede yapılacak olan iptaller, yüksek gelir kaybına neden olabilmektedir. Ve hizmete yakın zamanda yapılan bu iptaller geliri doğrudan ve olumsuz bir şekilde etkilemektedir. Yapılan bazı araştırmalara göre yapılan iptaller gelir kaybının yaklaşık %20 sini oluşturmaktadır. Bu durumdan kaynaklı olarak organizasyon ve planlama esas olmakla birlikte doğru bir tahmin aracına sahip olmak oldukça önemlidir. Rezervasyonların iptallerinden kaynaklanan gelir kaybının önceden tahmin edilebiliyor olması oldukça önemlidir

Turizm ve seyahatle ilgili sektörlerde gelir yönetimi, talep tahmini ve bu sorunların üzerine yapılan araştırmaların çoğu Yolcu Adı Kaydı (PNR) olarak bilinen formatta veriler kullanılmaktadır. Bu format havacılık endüstrisinin geliştirmiş olduğu bir formattır. Ancak konaklama, seyir vb. turizm ve seyahat endüstrileri keşfedilemeyecek farklı gereksinimlere ve özelliklere sahiptir. Bu araştırma kapsamında da talep verileriyle iki otel veri kümesi paylaşılmıştır. Bu çalışmada bir otel verisi kullanılarak iptal tahmini yapılmıştır.

Öncelikle turizm ve otel talep tahmin ve iptalleri ile ilgili bazı çalışmalar bulunmaktadır. Sanchez vd (2020) tarafından yapılmış olan çalışmada PNR verileri aracılığıyla Yapay Zeka teknikleri kullanılarak kısa bir zaman diliminde iptal etme olasılığı bulunan kişileri belirlemek amaçlanmıştır. Ve 7 gün öncesinden yapılan iptallerde %80 doğruluk değerini elde etmişlerdir.

Bu alanda daha önce yapılan çalışmalar nelerdir?

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

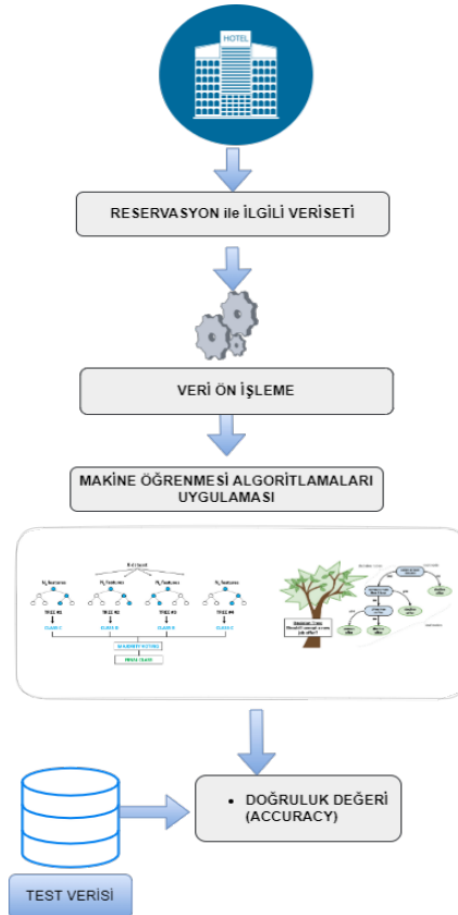
popüler olduğunu belirtmişler, özellikle Destek Vektör Makinesi'nin bu amaçla kullanılan en dikkate değer yöntem olduğunu belirtmişlerdir. Ayrıca rezervasyon hizmet zamanına ne kadar yakınsa, test edilen her teknikte o kadar az hata bulunması sonucuna da varmışlardır.

Bu çalışmada sınıflandırma algoritmalarından 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 depozito tipi ve müşterinin daha önce rezervasyon iptali yapıp yapmadığı sonucu elde edilmiştir.

Bu çalışma üzerinde neler yapılmış?

Yapılan bu çalışma otel rezervasyonlarından gelen veriler kullanılmaktadır ve sınıflandırma algoritmalarının yardımıyla rezervasyon iptalleri ve iptal edilmeme durumlarını doğru bir şekilde tahmin edebilen bir model oluşturmaktadır. Bu sınıflandırmada, bir otel rezervasyonu Normal veya İptal edildi olarak kategorize etmektedir. Bu çalışma müşterilerin rezervasyonunu rastgele iptal etmediğini varsaymaktadır. Bu çalışmada rezervasyon durumlarını doğru bir şekilde tahmin edebilmek için Karar Ağacı ve Random Forest Algoritmaları kullanılmıştır.

Bu çalışma kapsamında iptal tahmini için veriler erişilebilir web sitelerinden elde edilmiştir. İlk olarak veriler bir ön işlemden geçer ve bu aşamada veriler kodlanır. Analiz yapmak için veriler CSV formatına gönderilir. Veriler normalize edildikten sonra verilere Karar Ağaçları ve Random Forest algoritmaları uygulanır. Daha sonrasında test verisinde bir accuracy değeri oluşturulur.



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)

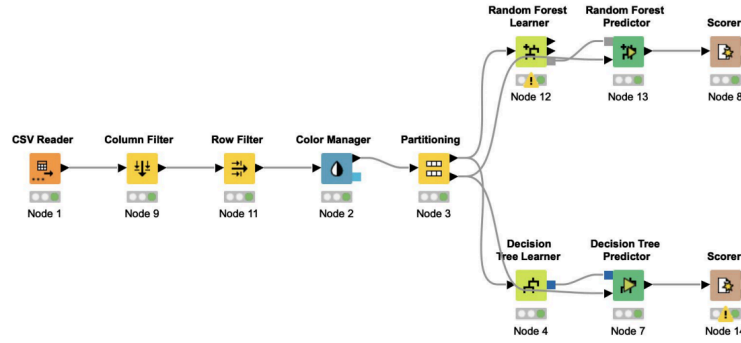
Bu çalışmada kullanılan veri kümesi incelendiğinde 28 farklı nitelik içerdiği gözlemlenmektedir.

Bu niteliklere bakıldığında, hedef nitelik Rezervasyon Durumudur. Rezervasyon Durumu , check-out, iptal veya no-show değerini alabilir.

Çalışmada kullanılan veri kümesi yaklaşık 80 bin rezervasyon içermektedir ve bu verilerin %41 i iptal ile sonuçlanmıştır.

Bu çalışmada çeşitli veri madenciliği metotları kullanılmıştır. Rezervasyon durumlarının doğru tahmin etme sonuçları karşılaştırılmıştır. Bu çalışma kapsamında Knime platformu üzerinde 2 farklı veri madenciliği metodu kullanılmıştır. (Karar Ağacı ve Random Forest).

Bu iki metod için de bir öğrenici ve bir de tahmin edici paket kullanılmıştır ve sonuçlar karşılaştırılmıştır.

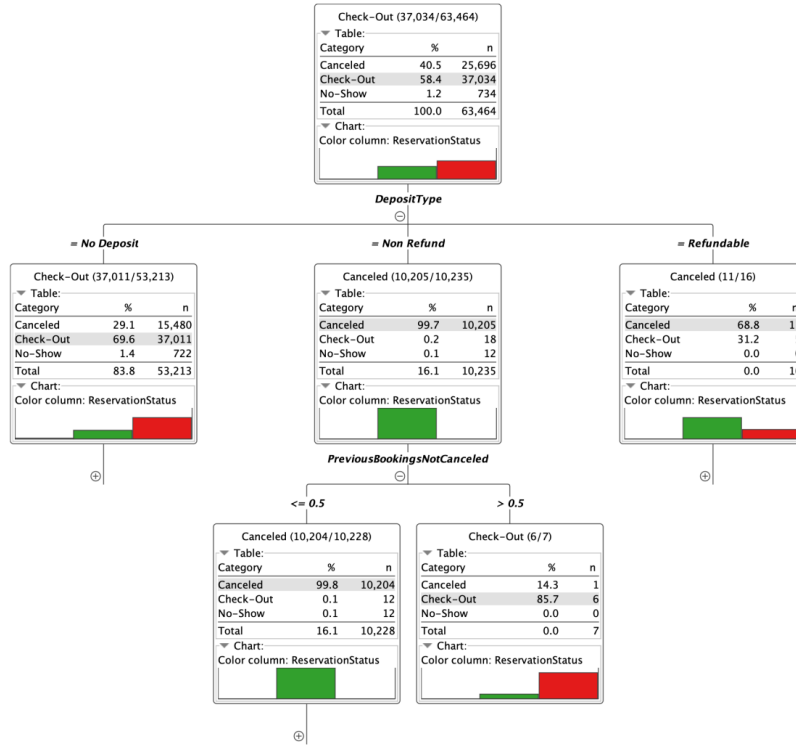


Şekil 3. Knime platformunda oluşturulan iş akışı

Bu iş akışı incelendiğinde ilk olarak veri bu platforma CSV formatında yüklenir. CSV dosyasındaki boş olan satır ve sütunlar temizlenir. Veri iki kümeye ayrılır. Bir küme eğitim verisi olarak belirlenirken diğer veri doğrulama verisi olarak kullanılır. Verinin %80 i eğitim kümesine ayrılırken, verilerin %20 si doğrulama amaçlı kullanılmıştır. Bu kümeyi bölme işlemi rastgele gerçekleştirilmiştir.

İlk olarak eğitim kümesiyle Karar Ağacı metodunun oluşturulduğu ağaç yapısı verilmiştir. Bu metodun sonucu incelendiğinde Rezervasyon durumu değişkenini en çok etkileyen 2 nitelik Depozito tipi ve müşterinin daha önce rezervasyon iptali yapıp yapmadığı sonucuna ulaşılmıştır.

Örneğin depozitosunu geri ödemeli olacak şekilde vererek rezervasyon yapan bir müşteri daha önce rezervasyon iptali yaptıysa bu rezervasyonunu da iptal etme olasılığı %99,8 olarak gözlemlenmiştir.



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.

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, Rastgele Orman modeli rezervasyon iptal tahmininde bariz bir avantaja sahiptir ve %86.7 doğruluk oranı ile otel rezervasyon tahmininde en yüksek sınıflandırma performansını elde etmektedir

MAKALE: <https://dergipark.org.tr/en/download/article-file/2607205>

Deniz Uzun
190301015