



Açıklanabilir Yapay Zekaya Dayalı Müşteri Kaybı Analizi ve Elde Tutma Önerisi Explainable AI-Driven Churn Analysis And Retention Recommendation

İbrahim KÖK 

¹Pamukkale Üniversitesi, Mühendislik Fakültesi, Pamukkale/Denizli, Türkiye

ikok@pau.edu.tr

Araştırma Makalesi/Research Article

ARTICLE INFO

Article history

Received : 16 August 2023

Accepted : 23 October 2023

Keywords:

Churn Analysis, Machine Learning, Explainable Artificial Intelligence, Subscriber Retention

ABSTRACT

In the mobile telecommunications market, subscribers expect high service quality, competitive pricing and improved service. If the customer cannot meet these expectations from the telecom service provider, he chooses to change it. To cope with this situation, service provider operators need interpretable churn prediction models that provide strategic insights by analyzing data on subscribers' communication patterns, behaviors and subscription plans. In this study, churn prediction models based on K-Nearest Neighbour, Decision Tree, Random Forest, Support Vector Machine and Naïve Bayes algorithms are first developed. Then, explainable artificial intelligence methods (ELI5, LIME, SHAP and Counterfactual) are used for the interpretability of the decisions of the most successful model. The experimental results show that the most successful model for subscriber churn detection is Random Forest with an accuracy of 81%. In addition, graphical results were obtained to explain how this algorithm makes decisions according to which features, how and with what accuracy rate. As a result, this study presents an application example of how to make both highly accurate and interpretable decisions in customer churn detection.

© 2024 Bandırma Onyedi Eylül University, Faculty of Engineering and Natural Science. Published by Dergi Park. All rights reserved.

MAKALE BİLGİSİ

Makale Tarihleri

Gönderim : 16 Ağustos 2023

Kabul : 23 Ekim 2023

Anahtar Kelimeler:

Müşteri Kaybı Analizi, Makine Öğrenmesi, Açıklanabilir Yapay Zeka, Abone Elde Tutma

ÖZET

Mobil telekomünikasyon pazarında aboneler yüksek hizmet kalitesi, rekabetçi fiyatlandırma ve gelişmiş servis beklentisindedirler. Müşteri bu beklentilerini telekom servis sağlayıcısından karşılayamaması durumunda onu değiştirme yoluna gitmektedir. Hizmet sağlayıcı operatörler abone kaybı olarak nitelendirilen bu durumla başa çıkmak için abonelerin iletişim kalıpları, davranışları ve abonelik planlarına ait verileri analiz ederek stratejik öngörü sağlayan yorumlanabilir müşteri kaybı tahmin modellerine ihtiyaç duymaktadırlar. Bu çalışmada öncelikle K-En Yakın Komşu, Karar Ağacı, Rastgele Orman, Destek Vektör Makinesi ve Naïve Bayes algoritmalarına dayalı abone kayıp tahmin modelleri geliştirilmiştir. Daha sonra, en başarılı modele ait kararların yorumlanabilirliği için açıklanabilir yapay zeka yöntemleri (ELI5, LIME, SHAP ve Counterfactual) kullanılmıştır. Deneysel sonuçlar abone kaybının tespitinde en başarılı modelin %81 doğruluk oranıyla Rastgele Orman olduğunu göstermiştir. Ayrıca bu algoritmanın kararları hangi özelliklere göre, nasıl ve ne kadar doğruluk oranıyla aldığına dair açıklamaları gösterir grafiksel sonuçlar çıkarılmıştır. Sonuç olarak, bu çalışmada müşteri kaybı tespitinde hem yüksek doğrulukla hem de yorumlanabilir kararların nasıl alınabileceğine ilişkin bir uygulama örneği sunulmuştur.

© 2024 Bandırma Onyedi Eylül Üniversitesi, Mühendislik ve Doğa Bilimleri Fakültesi. Dergi Park tarafından yayınlanmaktadır. Tüm Hakları Saklıdır.

1. GİRİŞ

Son yıllarda, nesnelerin interneti (IoT) ile birlikte mobil teknolojilerin hızla yayılması, dünya genelinde mobil cihaz sayısında büyük bir artışa neden olmuştur. Statista'nın tahminlerine göre, dünya çapında akıllı telefon mobil ağ abonelerinin sayısının 2023 yılında 6,7 milyara ulaşması ve 2028 yılında 7,7 milyarı aşması beklenmektedir [1]. Bu durum, telekomünikasyon sektöründeki servis sağlayıcılar için hem büyük bir fırsat hem de büyük bir zorluk oluşturmuştur. Sektör koşulları değiştiğinden, servis sağlayıcıların yeni müşteri çekme fırsatları artarken, mevcut müşterileri elde tutma zorluğunu aşmaları gerekmektedir. Ancak birçok şirket, doymuş pazarlardan kaynaklanan şiddetli rekabet, dinamik pazar koşulları ve sürekli yeni ticari tekliflerin sunulması nedeniyle önemli bir müşteri kaybı sorunu yaşamaktadır [2]. Dünyada telekomünikasyon sektörünün yıllık müşteri kayıp oranının %20 ila %40 arasında değiştiği ifade edilmektedir. Ayrıca mevcut müşterileri elde tutma maliyetinin yeni müşteri edinme maliyetinden 5-10 kat daha ucuz olduğu vurgulanmaktadır. Bunun yanı sıra müşteri kaybının %5 azaltıldığında kâr'ın %25 ten %85 çıkabildiği ortaya konmuştur [3]. Ayrıca, European Business Review [4] platformu da benzer bir görüşü paylaşmakta ve %2 oranında bir müşteri kayıp oranına sahip bir işletmenin aylık 65 milyon dolara kadar kayıplar yaşayabileceğini öngörmektedir. Bu durum pazar payı oldukça artan telekomünikasyon endüstrisinde rekabet artışını beraberinde getirmiştir [5]. Hizmet sağlayıcı operatörler rekabetçi ve doyuma ulaşmış pazar şartlarında yeni abone edinmenin maliyetinin mevcut bir aboneyi elde tutmadan daha fazla olduğunun farkındadırlar. Bu sebeple operatörler daha iyi servis kalitesi, ucuz tarifeler ve promosyonlarla abone kaybını önleme ve elde tutma stratejisini ön planda tutmaktadır [6]. Ancak abone verilerinin büyük olması, abonelerin operatörde ayrılma nedenlerinin çok değişken olması ve teknolojik gereksinimlerin hızla artması nedeniyle etkin bir müşteri kayıp modeli ve strateji oluşturmak oldukça zordur [7]. Bu ekonomik etki, büyük veriyi yönetme ve analiz etme yeteneği, dinamik değişken pazar koşullarına uyum sağlama konusunda hayati önem taşımaktadır. Veri analizi temelli çözümler, servis sağlayıcılara yeni veri kaynaklarından yararlanma ve müşteri davranışlarını daha derinlemesine anlama fırsatı sunacaktır. Bu veri analizi sayesinde operatörler, müşterilerin ihtiyaçları ve tercihleri konusunda daha öz ve anlamlı bilgileri hızlı şekilde erişebileceklerdir. Ayrıca, müşteri deneyimini iyileştirmek için özelleştirilmiş hizmetler sunma kapasiteleri artırma ve müşteri sadakati sağlama konusundaki yeteneklerini güçlendirme imkanı edineceklerdir [8]. Veri analizi, müşterilerin hangi hizmetlere daha fazla ilgi gösterdiklerini, hangi zamanlarda daha fazla kullanım yaptıklarını ve ne tür kampanyalardan daha fazla etkilendiklerini belirlemelerine yardımcı olacaktır. Bu bağlamda bir diğer önemli husus ise veri analiz sonuçlarının operatör politikası ve kampanyasına dönüşebilmesi için sonuçlarının operatörler tarafından kolaylıkla anlaşılabilir ve yorumlanabilir olması gerekmektedir.

Günümüzde yapay zeka alanının içinde yer alan ML (Machine Learning-ML) ve derin öğrenme algoritmaları mevcut alandaki bir çok sınıflandırma, kümeleme ve tahmin probleminde başarılı sonuçlar elde etmektedir. Ancak bu modeller çoğunlukla kara kutu modeller olarak değerlendirilmektedir [9]. Bu nedenle genellikle model önyargısı, model yapısı ve sonuçlarına ilişkin güven sorunlarına sahiptirler. Mevcut sorunlara çözüm olarak açıklanabilir yapay zeka (Explainable Artificial Intelligence-XAI) alanı geliştirilen yapay zeka sisteminin yüksek öğrenme performansını koruyarak daha açıklanabilir modeller üretmeye odaklanmaktadır. XAI ayrıca paydaşların geliştirilen modeli veya sistemi anlamasını, uygun şekilde güvenmesini ve etkili şekilde yönetmesini sağlayan şeffaf yöntemler çerçevesi sunmaktadır [10]. Bu sayede otomatik karar vermenin olumsuz sonuçlarıyla mücadele etmeleri için bireyleri güçlendirmek, bireylere daha bilinçli seçimler yapmalarında yardımcı olmak, güvenlik açıklarını ortaya çıkarmak ve algoritmaların insani değerlerle bütünleşmek hedeflenmektedir [9, 11].

Bu çalışmada literatürün bu eksikliğine yönelik hem ML algoritmalarına hem de açıklanabilir yapay zeka tekniklerine dayalı olarak etkili çözüm sunmaktadır. Bu bağlamda bu makalenin ana katkıları aşağıda sunulmuştur.

- Telekomünikasyon sektöründe kritik bir konu olan müşteri kaybı probleminde ML modellerine dayalı çözümler sunulmuştur. Bu sayede müşteri kayıp tahmini yüksek doğruluk oranlarıyla tahmin eden k-en yakın komşu, destek vektör makinesi, karar ağacı, rastgele orman ve naive bayes modelleri geliştirilmiştir.
- Geliştirilen ML modellerin etkili şekilde yorumlanabilmesi ve sonuçlarının anlaşılabilirliği için ELI5, LIME, SHAP ve Counterfactual gibi açıklanabilir yapay zeka yöntemleri kullanılmıştır.
- Ek olarak hizmet sağlayıcı operatörlere müşteri kaybını önleyici karşılıksal açıklanabilirlik yöntemine dayalı öneriler sunulmuştur.

Bu çalışmanın ilk bölümünde, mobil telekomünikasyon sektöründe müşteri davranış değişikliği ve müşteri kayıpları hakkında bilgiler verilerek konuya ilişkin literatür bilgisi aktarılmıştır. İkinci bölümde, Telekomünikasyon sektöründe müşteri kaybı tahmine yönelik yapılan çalışmaları ait literatür bilgisi verilmiştir. Üçüncü bölümde, bu çalışmada kullanılan materyal ve metotlara ilişkin teorik bilgiler sunulmuştur. Dördüncü bölümde, deneysel sonuçlar yer verilmiş ve elde edilen uygulama sonuçları karşılaştırmalı olarak değerlendirilmiştir. Dördüncü bölümde ise çalışma sonuçlandırılmıştır.

2. İLGİLİ ÇALIŞMALAR

Bu çalışma kapsamında ise telekomünikasyon sektöründe müşteri kaybının tahminine odaklanan güncel literatür çalışmalarına yer verilmiştir.

Kişioğlu ve Topçu [12], Türkiye'deki bir telekomünikasyon şirketi verilerini kullanarak Bayes İnanç Ağına (Bayesian Belief Network) dayalı müşteri kayıp modeli oluşturmuşlardır. Yazarlar modelin sonuçlarını daha anlaşılır hale getirmek için ise ki-kare otomatik etkileşim detektörü, korelasyon analizi, çoklu bağlantı testi ve uzman görüşü gibi yöntemleri ek olarak kullanmışlardır. Müşteri kaybına neden olan değişkenleri ortalama görüşme dakikaları, ortalama fatura tutarı, farklı sağlayıcılardan kişilere arama sıklığı ve tarife türü olarak tespit etmişlerdir. Lu ve arkadaşları [13] ise müşteri kaybı tahmini için öncelikle artırma (boosting) yöntemine dayalı ağırlıklandırma yöntemi kullanmışlardır. Daha sonra logistic regresyon modeline dayalı müşteri kayıp tahmini gerçekleştirmişlerdir. Jung ve arkadaşları [14] veriye dayalı müşteri deyim kalitesini artırarak müşteri kaybını önlemeyi amaçlamışlardır. Özellikle müşteri şikayetleri olasılığını tahmin etmek için rastgele orman, destek vektör makinesi, lasso ve karar ağacı algoritmalarını kullanmışlardır. Bagdi ve arkadaşları [15] telekomünikasyon veri üzerinde destek vektör makinesi ve rastgele orman algoritmasına dayalı müşteri kayıp modelleri geliştirmişlerdir. Geliştirilen modellerin sonuçlarını anlamlandırmak için ise k-ortalama kümeleme, histogram, kutu ve bar grafikleri oluşturmuşlardır. Mitrovic ve arkadaşları [16] müşteri kaybının tespitinde özellik mühendisliğinin önemine dikkat çekmişlerdir. Bu kapsamda kayıp analizinde optimum özellik türü kombinasyonlarını belirlemek için Pareto çok amaçlı optimizasyona dayalı yeni bir metodoloji önermişlerdir. Önerilen metodu kullanarak belirlenen özelliklerle rastgele orman ve lojistik regresyon modeli ile sınıflandırma gerçekleştirmişlerdir. Önerilen metod ve sınıflandırma sonuçlarının müşteri kaybı modellenmesinde kılavuz olabileceğini ortaya koymuşlardır. Caigny ve arkadaşları [17] birçok sektörün temel problemi olan müşteri kaybı tahmini için karar ağaçları ve lojistik regresyon modeline dayalı hibrit logit yaprak modeli (LLM) önermişlerdir. Önerilen modelin temel fikri tüm veri kümesi yerine ağacın her yaprağı üzerindeki veri parçaları üzerinde inşa edilen farklı modeller aracılığıyla tahmin yapmasıdır. Önerilen model finans, enerji, telekom, perakende ve gazete gibi sektörlerden 14 veri seti üzerinde test edilmiş ve karar ağaçları, lojistik regresyon ve rastgele orman modellerinden daha iyi olduğu gösterilmiştir. Cenggoro ve arkadaşları [18] ise müşteri kaybını modellemek için ayırt edici özelliklerin belirlenmesinde gömülü vektörler (vector embedding) modeline dayalı derin öğrenme modellerinin geliştirilmesine odaklanmışlardır. Yazarlar önerdikleri modelle vektörlerin iki boyutlu bir uzayda görselleştirilerek ayrılma ve elde tutulma potansiyeli yüksek olan iki grubu etkili şekilde tespit edebilmişlerdir.

Wu ve arkadaşları [19] telekomünikasyon alanında müşteri kayıp yönetimi için veri ön işleme, keşifsel veri analizi, kayıp tahmini, faktör analizi, müşteri segmentasyonu ve müşteri davranışı analitiği gibi altı bileşeni içeren bir çerçeve önermişlerdir. Önerilen çerçevede müşteri kayıp tahmini için karar ağaçları, rastgele orman, naive bayes, Adaboost ve çok katmanlı algılayıcı modelleri kullanmışlardır. Model sonuçlarına dayalı müşteri kaybının nedenlerinin tespit için faktör analizi ve kümeleme yöntemlerini kullanmışlardır. Shrestha ve Shakya [20] müşteri kaybı tahmini için XGBoost algoritmasının etkinliğini analiz etmeye odaklanmıştır. İki farklı veri seti üzerinde elde edilen sonuçları literatürdeki diğer çalışmalarla karşılaştırmışlardır. Deneysel sonuçlarda %97 doğruluk oranı ile XGBoost algoritmasının müşteri kaybı tahmin problemleri için diğer sınıflandırıcılara göre daha iyi sonuçlar üretebildiğini göstermişlerdir. Amin ve arkadaşları [21] telekomünikasyon sektöründe müşteri kayıp tahmini için uyarlanabilir ML'ye dayalı ACCP adında yeni bir yaklaşım önermişlerdir. Yazarlar önerdikleri yaklaşımda genetik algoritma tabanlı özellik ağırlıklandırma yaklaşımı ile naive bayes algoritmasını kullanarak hibrit bir yaklaşım ortaya koymuşlardır. Önerilen yaklaşım sinir ağları, lineer regresyon, KNN, XGBoost gibi temel sınıflandırıcılarla karşılaştırmışlardır. Bir diğer çalışmada, Saha ve arkadaşları [22] telekom sektöründe müşteri kayıp tahmini için geleneksel sınıflandırma teknikleri (rastgele orman, xgboost, gradyan artırma), topluluk öğrenme teknikleri (lojistik regresyon, karar ağacı, k-en yakın komşu) ve derin öğrenme (yapay sinir ağları) modellerine dayalı kapsamlı çalışma yapmışlardır. Müşteri kayıp tahmininde yapay sinir ağları ve konvolüsyonel sinir ağlarının yaklaşık %98 doğruluk oranıyla diğer modellerden daha iyi olduğunu göstermişlerdir. Prabhadavi ve arkadaşları [23] telekomünikasyon sektöründe müşteri kaybını erken tahmin etmek için optimum ML modelleri geliştirmeyi hedeflemişlerdir. Bu kapsamda mevcut müşterileri elde tutmak için gösterecekleri tepkileri tahmin etmek için stokastik gradyan güçlendirici, rastgele orman, lojistik regresyon ve k-en yakın komşu algoritmalarını kullanmışlardır.

Ayrıca literatürde bulanık mantık ve teorik modellere dayalı analiz yöntemlerinin de kullanıldığı bazı çalışmalar da bulunmaktadır [24-27]. Tüm literatür incelendiğinde müşteri kaybı analizi ve tahmininde ML, derin öğrenme, veri madenciliği ve istatistiksel yöntemleri dayalı birçok modelin kullanıldığı açıkça görülmektedir. Ancak literatürde yer alan bu modellerin açıklanabilirliğine ve yorumlanabilirliği odaklanan bir çalışma bulunmamaktadır. Özellikle açıklanabilir yapay zeka destekli modellerin olmayışı ele alınan probleme yönelik geliştirilen modelleri anlama ve onlara güven konusunda problem yaratmaktadır. Bazı çalışmalarda açıklanabilirlik için birden çok istatistiksel yöntem (ki-kare, faktör analizi, kümeleme gibi) tercih edilmiş olsa da bunlar alanda ihtiyaç duyulan gereksinimleri karşılamada yetersiz kalmaktadır. Bu nedenle bu çalışmada müşteri kaybı tahmininin açıklanabilir yapay zeka yöntemlerine dayalı çözüm önerisine odaklanılmıştır.

3. MATERYAL VE METOD

Bu bölümde bu makalede kullanılan veriseti, veri ön işleme süreci, ML algoritmaları ve açıklanabilir yapay zeka yöntemlerine yönelik detaylı bilgiler verilmektedir.

3.1. Veri Seti

Bu çalışmada IBM telekomünikasyon şirketine ait 20 özellik ve 7043 satırdan oluşan Telco müşteri veri seti [28] kullanılmıştır. Veri setinde abonelere ait demografik bilgiler, hesap bilgileri ve hizmet bilgileri gibi özellikler yer almaktadır. Veri setinde yer alan özellikler ve açıklamalarına ait kapsamlı bilgiler Tablo 1’de verilmiştir.

Tablo 1. Telco veriseti özellikleri ve açıklamaları.

Özellik	Açıklama	İçerik
Gender	Cinsiyet	Kadın, Erkek (Female, Male)
SeniorCitizen	Müşterinin kıdemli/emekli/yaşlı olup olmadığı	Evet, Hayır (Yes, No)
Partner	Müşterinin bir ortağı olup olmadığı	Evet, Hayır (Yes, No)
Dependents	Bakmakla yükümlü olduğu kişi olup olmadığı	Evet, Hayır (Yes, No)
Tenure	Müşterinin şirkette kaldığı ay sayısı	Nümerik değer
PhoneService	Müşterinin telefon hizmeti olup olmadığı	Evet, Hayır (Yes, No)
MultipleLines	Müşterinin birden fazla hattı olup olmadığı	Evet, Hayır, Telefon Servisi Yok (Yes, No, No phone service)
InternetService	Müşterinin internet servis sağlayıcısı	DSL, Fiber optik, Hayır (DSL, Fiber optic, No)
OnlineSecurity	Müşterinin çevrimiçi güvenliğinin olup olmadığı	Evet, Hayır, Internet Servisi Yok (Yes, No, No Internet service)
OnlineBackup	Müşterinin online yedeğinin olup olmadığı	Evet, Hayır, Internet Servisi Yok (Yes, No, No Internet service)
DeviceProtection	Müşterinin cihaz korumasına sahip olup olmadığı	Evet, Hayır, Internet Servisi Yok (Yes, No, No Internet service)
TechSupport	Müşterinin teknik destek alıp almadığı	Evet, Hayır, Internet Servisi Yok (Yes, No, No Internet service)
StreamingTV	Müşterinin TV yayını olup olmadığı	Evet, Hayır, Internet Servisi Yok (Yes, No, No Internet service)
StreamingMovies	Müşterinin canlı yayım/film akışı olup olmadığı	Evet, Hayır, Internet Servisi Yok (Yes, No, No Internet service)
Contract	Müşterinin sözleşme süresi	Aydan aya, Bir yıl, İki Yıl (Month-to-month, One year, Two year)
PaperlessBilling	Müşterinin kağıtsız faturası olup olmadığı	Evet, Hayır (Yes, No)
PaymentMethod	Müşterinin ödeme yöntemi	Elektronik çek, Posta çeki, Banka transferi, Kredi Kartı (Otomatik)
MonthlyCharges	Müşteriden aylık olarak tahsil edilen tutar	Nümerik değer
TotalCharges	Müşteriden tahsil edilen toplam tutar	Nümerik değer
Churn	Abone hareketi - Müşteri kaybı	Evet, Hayır (Yes=1, No=0)

3.1. Veri Önleme

Bu çalışmada kullanılan Telco veriseti ML modelleriyle eğitilmeden önce gereksiz verilerin temizlenmesi ve veri dönüşümü (label encoding) süreçlerinden geçirilerek model eğitimine hazırlanmıştır. Ayrıca verisetinin dengesiz olması (yaklaşık %74 churn no, %26 churn yes) nedeniyle eğitim sürecinde veriyi dengeli hale getirmek için örneklem artırma yöntemi SMOTE (Synthetic Minority Over-sampling Technique) kullanılmıştır. Bu bağlamda dengesiz veri probleminin sonuca etkisini analiz edebilmek için sınıflandırma sonuçları hem dengeli hem de dengesiz veri seti dikkate alınarak ayrı ayrı verilmiştir.

3.2. Makine Öğrenmesi Algoritmaları

Bu çalışmada odaklanılan müşteri kaybı tahmini için literatürde daha çok ML algoritmalarına yönelik sınıflandırma yaklaşımların kullanıldığı görülmüştür. Bu nedenle bu çalışmada ilgili problemin çözümünde ML modellerine ek olarak açıklanabilir yapay zeka yöntemleri de kullanılmıştır. Bu bölümde öncelikle kullanılan ML modelleri açıklanmıştır. Bu modeller sırasıyla K-En Yakın Komşu (K-Nearest Neighbor-KNN), Karar Ağacı (Decision Tree-DT), Rastgele Orman (Random Forest-RF), Destek Vektör Makinesi (Support Vector Machine-SVM) ve Naïve Bayes (NB) algoritmalarıdır. Bu algoritmalara ait özet açıklamalar aşağıda sunulmaktadır.

3.2.1. K-En Yakın Komşu (K-Nearest Neighbor-KNN)

KNN algoritması örnek veri noktasını tahmin etmek için özellik benzerliğini kullanan tembel ve non-parametrik bir öğrenme algoritmasıdır. Algoritmanın temel mantığı veriler arası uzaklığa dayanmaktadır. Bir verinin hangi sınıfa dahil olduğuna karar verebilmek için o veriye en yakın “k” adet veriye bakılmaktadır. En çok hangi sınıfa ait veri varsa test verisi de o sınıfa ait kabul edilmektedir. Bu algoritmanın başarısı “k” değerinin seçimine ve uzaklık hesabı için hangi metodun kullanılacağına bağlıdır. Komşuluk sayısı çok az ya da çok fazla seçilirse doğruluk oranı düşebilir. Bu yüzden farklı komşuluk sayıları ile algoritmanın tekrarlanması ve en iyi sonucu veren “k” değerinin bulunması gerekmektedir. Bu algoritmada kullanılan uzaklık metriğinin hesaplanmasında birden

farklı yöntem bulunmaktadır. Bunlardan en çok bilineni öklid uzaklığıdır. Denetimli ML algoritması olup sınıflandırma ve regresyon problemlerinde yaygın olarak kullanılmaktadır [29].

3.2.2. Karar Ağacı (Decision Tree-DT)

Karar ağaçları, bağımlı değişken üzerindeki farklılıkların maksimize edilmesi amacıyla veri setinin belirli karar adımlarına göre bölünerek oluşturulan yapılardır. Ağaçtaki her bir nitelik bir düğüm tarafından temsil edilmektedir. Dallar ve yapraklar ağaç yapısının elemanlarıdır. En son yapı *yaprak*, en üst yapı *kök* ve bunların arasında kalan yapılar ise *dal* olarak adlandırılır. Ağaç oluşturulurken dallanma kriteri olarak entropi'ye dayalı, sınıflandırma ve regresyon ağaçları ve bellek tabanlı yaklaşımlar bulunur. Anlaşılır kurallar üretilebilmesi, daha az hesaplama ve sınıflama gerektirmesi, sürekli ve kategorik verilerde kullanılabilmesi karar ağaçlarının avantajları arasında sayılmaktadır. Ancak karar ağaçları eğitim aşamasında hesaplama açısından masraflı olabilir. Sürekli değerleri tahmin için az elverişli olması ve küçük eğitim örneklerinde çeşitli sınıflama problemleri ise dezavantajları olarak verilmektedir [29].

3.2.3. Rastgele Orman (Random Forest-RF)

RF algoritması birden fazla karar ağacını bir araya getirerek daha güvenilir ve doğru sonuçlar üretmeye çalışan topluluk öğrenme algoritmasıdır. Bu algoritmada her bir karar ağacı, veri kümesinin rastgele bir alt kümesi üzerinde eğitilmekte ve girdi verilerine dayalı olarak birçok karar kuralı oluşturulmaktadır. Ardından, bu ağaçlar tahminlerde bulunmakta ve sonuçlar birleştirilmektedir. RF, karar ağaçlarının bir araya gelerek daha güvenilir tahminler yapmasını sağlayarak overfitting riskini azaltmaktadır. Bu nedenle, genellikle büyük veri setlerinde ve karmaşık veri yapılarında tercih edilen bir algoritmadır [30]. RF özellikle sınıflandırma ve regresyon problemleri yaygın olarak kullanılmaktadır.

3.2.4. Destek Vektör Makinesi (Support Vector Machine-SVM)

SVM, verileri yüksek boyutlu uzayda temsil ederek bu verileri ayıran hiper düzlemi (karar sınırı) bulmaya çalışan bir algoritmadır. Hiper düzlemi sınıflar arasındaki marjini maksimize ederek ve destek vektörleri belirleyip onları optimize ederek oluşturmaktadır. SVM, doğrusal, polinom, radyal temel işlev ve sigmoid gibi çekirdekler (kernellere) yardımıyla lineer ve non-lineer sınıflandırmaları etkili şekilde gerçekleştirebilmektedir. Hem çalışma mantığı, yapısı ve performansı açısından özellikle düşük veri durumlarında yüksek hassasiyetle çalışabilmektedir [30].

3.2.5. Naïve Bayes (NB)

NB algoritması bayes teoremine dayalı olarak sonrasal olasılıkları hesaplamakta kullanılan ve rastgele iki olayın koşullu ve marjinal(önsel) olasılıklarını ilişkilendiren istatistiksel sınıflandırma algoritmasıdır. Bu algoritmadaki temel varsayım, veri setindeki özelliklerin koşullu olarak bağımsız olduğudur. Bu durum çok sayıda özelliği işlemesine ve sınırlı eğitim verilerinde iyi çalışmasına imkan sağlamaktadır. Bu algoritmada bir olayın koşullu olasılığını hesaplanırken bu olayın gerçekleştiği bir başka olayın etkisi dikkate alınmaktadır. Örneğin, $P(A|B)$ hesaplaması, $P(B|A)$ ve bağımsız olarak $P(A)$ ve $P(B)$ bilgilerini kullanarak yapılmaktadır.

Bu hesaplama Denklem 1'deki bayes teoremine dayalı olarak gerçekleştirilmektedir. Bu teorem, özellikle makine öğrenimi ve istatistiksel analiz gibi alanlarda, belirsizlik altında olayların olasılıklarını tahmin etmek için yaygın olarak kullanılmaktadır [31].

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \quad (1)$$

$P(A|B)$; B olayı gerçekleştiği durumda A olayının meydana gelme olasılığını, $P(B|A)$; A olayı gerçekleştiği durumda B olayının meydana gelme olasılığını, $P(A)$ ve $P(B)$; A ve B olaylarının önsel olasılıklarını ifade etmektedir.

3.3. Açıklanabilir Yapay Zeka (Explainable Artificial Intelligence-XAI) Yöntemleri

Yapay Zeka (AI) modellerinin kara kutu yapısı, kullanıcıların bu model tarafından oluşturulan çıktıyı anlamalarına ve bazen bunlara güvenmelerine izin vermez. Yalnızca sonuçların değil sonuçlara giden karar yollarının da kritik olduğu AI uygulamalarında, bu tür kara kutu AI modellerinin kullanıcılar için daha anlaşılır, güvenilir ve şeffaf olması gerekmektedir. Bu noktada Açıklanabilir Yapay Zeka (XAI), AI tabanlı sistem bilgileri ve kararlarının insan tarafından anlaşılabilir ve yorumlanabilir açıklamalarını üreten yaklaşımları, algoritmaları ve araçları incelemekte ve tasarlamaktadır [32]. XAI yapay zeka girdi verilerinin, model ve model sonrası karar verme mekanizmalarının aydınlatılmasını veya ortaya çıkarılmasını amaçlamaktadır [11]. Bu amaca yönelik olarak veri ön işleme aşamasında model geliştirme sürecinde ve model sonuçlarının kullanıcı ve uzman tarafından anlaşılması sürecinde model sonrasında birçok XAI yöntemi kullanılmaktadır. XAI yöntemlerinin temel amacı gelişmiş gerekçelendirme, kontrol, iyileştirme ve keşiftir [9].

3.3.1. Yerel Yorumlanabilir Model-agnostik Açıklamalar (Local Interpretable Model-agnostic Explanations-LIME)

Model kararlarının yorumlanabilirliğine odaklanan modelden bağımsız bir yaklaşımdır. Her bir bireysel tahmini açıklamak için herhangi bir kara kutu ML modeline yerel, yorumlanabilir bir modelle yaklaşmayı amaçlamaktadır. Bu yaklaşımın temelinde girdileri bozarak ve bu durumun modelin çıktılarını nasıl etkilediğini izlemek vardır. Buna göre, örnek veri noktalarını bozarak yeni örnekler oluşturur ve bu örneklerin tahminlerini anlamak için basit bir açıklanabilir model oluşturur. Oluşturulan bu model, orijinal modelin kararlarına benzer tahminler yapar ve bu sayede tahminlerin nedenlerini daha iyi anlamamıza yardımcı olur. LIME, özellikle tıbbi teşhisler, kredi riski değerlendirmeleri gibi kritik uygulamalarda, model kararlarının anlaşılabilirliğini artırmak için etkili bir araç olarak kullanılabilir [33].

3.3.2. SHapley Katkı Açıklamaları (SHapley Additive exPlanations-SHAP)

SHAP yöntemi oyun teorisindeki Shapley değeri olarak bilinen matematiksel kavramı kullanarak özniteliklerin çıktıya katkısını bulmaya odaklanan XAI yöntemidir. Bu yöntemde kullanılan Shapley değeri, özellik uzayındaki tüm olası değerler arasında her bir özellik değerinin ortalama marjinal katkısını ifade etmektedir. SHAP hem global hem de yerel açıklamalarda kullanılabilir. Teorik olarak sağlamdır ancak yüksek boyutlu verilerde Shapley değerini hesaplamak maliyetli olduğunda kullanılması dezavantaj oluşturmaktadır [34].

3.3.3. Beş Yaşındaymışım Gibi Açıkla (Explain Like I'm 5 - ELI5)

Bu yöntem isminde de belirttiği üzere karmaşık ve teknik olarak zor konuları bir çocuğa anlatır gibi basit ve anlaşılabilir şekilde açıklama amacı taşıyan bir yöntemdir. Model parametreleri incelemeye ve modelin genel özellikleri çıkarmada etkili şekilde çalışmaktadır. Ayrıca global ve yerel açıklamalarla modelin tüm veriseti özelinde veya bireysel tahminlerini neden aldığına dair geçerli açıklamalar sağlamaya yardımcı olur [34].

3.3.4. Karşı olgusal açıklama (Counterfactual)

Karşı olgusal (karşıt gerçeklik) açıklama yöntemi bir örnek veri noktasının tahmin sonucunu değiştirmek için gerekli olan değişiklikleri gösteren ve çoklu öneriler sunan XAI yöntemidir. Model çıktısının tahminin tersi durumunda girdi tarafında hangi değişkenlerin ne kadar değişmesi gerektiğine odaklanmaktadır. Örneğin ikili sınıflandırma probleminde model çıktısı 0 olduğu durumda, çıktıyı 1 yapacak girdi değerlerini bizlere sunmaya çalışır. Her örnek uzayı için birden çok karşı olgusal çıktı üretebilmektedir. Burada zorluk karşı olgusal açıklamalardan hangisi olduğudur. Karşı olgusal yöntemler, verilere veya modellere erişim gerektirmez ve makine öğrenimini hiç kullanmayan bir sistemle de çalışabilmektedir [35].

4. DENEYSEL SONUÇLAR VE TARTIŞMALAR

Bu bölümde abone kaybı tahmini için geliştirilen ML algoritmaları ve açıklanabilir yapay zeka yöntemlerine ilişkin deneysel sonuçlar ve karşılaştırmalı değerlendirmeler her alt bölümün sonunda sunulmuştur.

4.1. Makine Öğrenmesi Algoritmalarına İlişkin Sonuçlar

Bu çalışmada kullanılan ML modellerinin performansı hem dengeli hem de dengesiz veriseti kullanılarak elde edilmiş ve Tablo 2'de verilmiştir. Çalışma kapsamında geliştirilen ML modellerinin tahmin performansını değerlendirmek için kesinlik, duyarlılık, F1-skoru ve doğruluk metrikleri kullanılmıştır. Ayrıca bu çalışmada kullanılan veri setini kullanan literatürdeki diğer çalışmalarla doğruluk metriği açısından karşılaştırma yapılarak Tablo 3'te sunulmuştur.

Tablo 2'deki SMOTE yöntemi kullanılmadan dengesiz veriseti (yaklaşık %74 churn no, %26 churn yes) ile yapılan sınıflandırma sonuçları değerlendirildiğinde en başarılı sonuçlar %81, %80, %80 doğruluk oranlarıyla sırasıyla RF, SVM ve DT algoritmaları tarafından elde edilmiştir. Bu sonuçlara göre abonelerin operatörde kalma ve operatörü terk etme durumu en başarılı belirleyen RF algoritması olmuştur. Bu sonuçta RF algoritmasının birden çok ağaç modelini baz alarak topluluk öğrenme yaklaşıma dayalı olarak çalışması etkili olmuştur.

Diğer taraftan SMOTE yöntemi ile dengeli hale getirilen veriseti sınıflandırma sonuçlarına göre sırasıyla en başarılı modeller %77, %76 ve %75 doğruluk oranlarıyla RF, SVM ve KNN olmuştur. Dengeli veri setinde sınıflandırma başarı oranlarının biraz daha düşük olduğu gözlemlenmektedir. Bu sonucun nedeni, dengesiz veri setlerinde baskın sınıfın daha fazla örneği olduğu için model, baskın sınıfı daha iyi tahmin edebilirken azınlık sınıfı daha zayıf tahmin edebilmektedir. Dengesizlik, azınlık sınıfın doğru tahmin edilmesini zorlaştırabilir. Dengeli bir veri seti, modelin tüm sınıfları daha dengeli bir şekilde tahmin etmesini sağlayabilmekte ve sonucun istatistiksel olarak daha stabil olmasını sağlamaktadır.

Bu çalışmada elde edilen sonuçlar literatürde aynı veri setini kullanan başka çalışmalarla da karşılaştırılarak sayısal sonuçlar Tablo 3'te verilmiştir. Buna göre [19] referans numaralı çalışmada hem dengesiz hem de dengeli veri setiyle yapılan sınıflandırmalarda en başarılı RF algoritması olmuş ve sırasıyla yaklaşık %79 ve %76 başarı oranları elde

Tablo 2. ML modelleri sınıflandırma sonuçları

Model	Dengesiz Veriseti (Without SMOTE)				Dengeli Veriseti (With SMOTE)			
	Kesinlik (Precision)	Duyarlılık (Recall)	F1-Skoru (F1-score)	Doğruluk (Accuracy)	Kesinlik (Precision)	Duyarlılık (Recall)	F1-Skoru (F1-score)	Doğruluk (Accuracy)
KNN	0.781108	0.794890	0.784278	0.794890	0.754581	0.758694	0.756506	0.758694
DT	0.787835	0.801278	0.790272	0.801278	0.776141	0.748048	0.758004	0.748048
RF	0.799160	0.811214	0.800758	0.811214	0.792074	0.772179	0.779533	0.772179
SVM	0.795421	0.804829	0.798427	0.804826	0.796871	0.765082	0.775375	0.765082
NB	0.782992	0.799148	0.774352	0.799148	0.778779	0.733144	0.747096	0.733144

edilmiştir. [23] referans numaralı çalışmada ise sadece dengesiz veri sonuçları alınmıştır ve en başarılı model yaklaşık %80 doğruluk oranıyla RF algoritması olmuştur. Bizim çalışma sonuçlarımız ise hem dengesiz hem de dengeli veri setinde literatürdeki diğer çalışmalardan daha başarılı sonuçlar ortaya koymuştur. Diğer taraftan hem bizim çalışmamızda hem de diğer çalışmalarda en başarılı algoritma olarak RF algoritması ön plana çıkmıştır.

Tablo 3. Bu çalışmanın literatürdeki çalışmalarla doğruluk metriği açısından karşılaştırılması

Model	Dengesiz Veriseti (Without SMOTE)			Dengeli Veriseti (With SMOTE)		
	Bu çalışma	Wu et al. [19]	Prabadevi et al. [23]	Bu çalışma	Wu et al. [19]	Prabadevi et al. [23]
KNN	0.794890	-	0.77730	0.758694	-	-
DT	0.801278	0.783	-	0.748048	0.767	-
RF	0.811214	0.795	0.80388	0.772179	0.769	-
SVM	0.804826	-	-	0.765082	-	-
NB	0.799148	-	-	0.733144	-	-

- (tire) : İlgili çalışmada gerekli veri bulunmamaktadır.

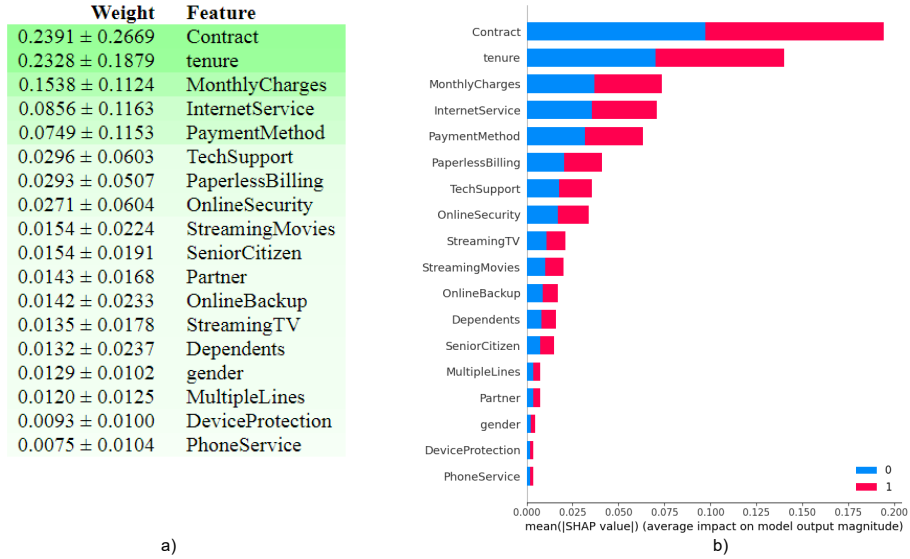
Yukarıdaki bilgiler ışığında literatürde müşteri kaybı tahmininde hem dengeli hem de dengesiz veriler kullanıldığı görülmektedir. Bu çalışmalarda sadece tahmin modeli geliştirilme sonuçları başarı oranları özelinde kıyaslanmaktadır. Bu noktada abonenin sadece operatöründe kalma durumu veya ayrılma durumu tespit etmek abonenin tutulmasına yönelik politika ve kampanya geliştirilmesi için yeterli değildir. Elde edilen sonucun hangi oranda, nasıl ve neden kaynaklandığını tespit etmek kritik önemdedir. Bu nedenle sonraki bölümde RF algoritması sınıflandırma sonuçlarının açıklanabilirliğine ilişkin çalışma sonuçlarına yer verilmiştir.

4.2. Açıklanabilir Yapay Zeka Yöntemleri İlişkin Sonuçlar

Bu bölümde geliştirilen ML modellerinin yorumlanabilirliği için ELI5, LIME ve SHAP yöntemlerini kullanıyoruz. Burada geliştirilen en başarılı sınıflandırma modeli olan RF'ye göre hem global hem de lokal açıklanabilirlik sonuçlarını verilmiştir. Bu kapsamda öncelikle RF modelinin tüm veri seti üzerindeki sınıflandırma kararına etki eden özellikler tespit etmek ve modelin davranışına bütüncül olarak bakmak için ELI5 yöntemiyle global açıklamalar çıkarılmıştır. Elde edilen global açıklamalara çıktılar Şekil 1'de verilmiştir.

Şekil 1'teki sonuçlar tüm müşteriler/veri seti bağlamında mobil operatör abonelerin kaybedilmesinde genel olarak etkili olan özellikleri önem sırasına göre vermektedir. Bu bilgilere göre abone kaybına neden olan ilk beş özelliğin sözleşme (contract) süresi, abonenin şirkette kaldığı ay sayısı (tenure), aylık ödeme tutarları (MonthlyCharges), internet servisleri (InternetService) ve ödeme yöntemleri olduğu görülmektedir. İlk iki özellik bağlamında abonelerin durumu detaylı incelendiğinde operatör değiştiren (y=1) abonelerin %88'inin aylık sözleşmeler kullandığı ve bir operatörde yaklaşık ortalama 17 ay kaldığı tespit edilmiştir. Diğer taraftan operatörüne sadık kalan abonelerin ise yaklaşık %57 sinin bir yıl ve üzeri sözleşmeye sahip olduğu ve operatöründe kalma süresi ortalama 38 ay olduğu tespit edilmiştir. Bu sonuçlardan sözleşme sürelerinin uzun ve aylık ücretlerin düşük tutulması aboneleri operatöre bağlı tutacağı yorumu çıkarılabilir. Bunların dışında internet servisleri, ödeme yöntemleri ve teknolojik destek gibi diğer bileşenlerinde abonelere sağlanması müşteri kaybını önlemede etkili olduğu sonucu çıkarılabilir.

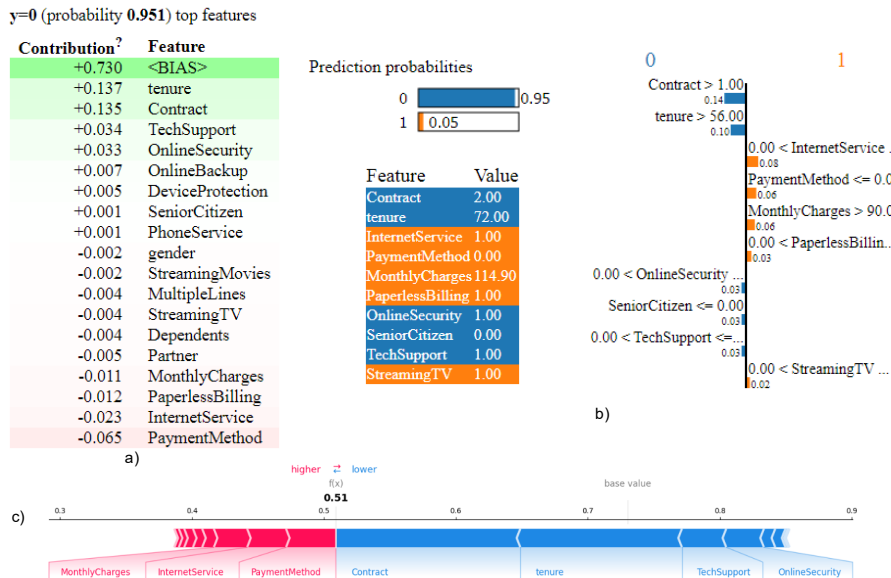
Global açıklamalar tüm aboneler bağlamında genel politikalar üretmede yardımcı olurlar. Ancak günümüzde aboneleri operatöre bağlamak için müşteri spesifik kampanya ve seçenekler sunmak kritik önemdedir. Bu nedenle her bir müşterinin tercihlerini anlamak ve ihtiyaçlarına yönelik açıklama getirmek gerekmektedir. Bu bağlamda veri setinden biri operatöre bağlı/ayrılmayan (y=0), diğeri operatörden ayrılan (y=1) olmak üzere iki örnek seçilmiştir. Seçilen bu iki örnek aboneye ait verilerin lokal açıklamaları Şekil 2 ve Şekil 3'te verilmiştir. Şekil 2'de RF modelimizin %95 olasılıkla sonucu 0 yani mobil operatörden ayrılmayan olarak tahmin ettiği bir abonenin ELI5, LIME ve SHAP yöntemlerine dayalı lokal açıklanabilirlik sonuçları yer almaktadır. Bu arada her üç yöntemin de benzer çıktılar ürettiği gözlemlenmiştir.



Şekil 1. RF modelinin ELI5 ve SHAP yöntemleriyle global açıklanabilirliği (a. ELI5 yöntemi global açıklama, b. SHAP yöntemi global açıklama).

Buna göre ilgili abonenin operatörde kalmasında sözleşme (contract) süresi, abonenin şirkette kaldığı ay sayısı (tenure), teknikdestek (TechSupport), online güvenlik (OnlineSecurity) gibi seçeneklerin öncelikle etkili olduğu görülmektedir. Detaylı veri analizinde bu abonenin iki yıllık sözleşmesi olduğu ve operatörde 72 aydır bulunduğu görülmektedir. Bu bağlamda müşterinin sözleşme süresine bağlı kaldığı ve verilen teknik hizmet ve güvenlikten memnun olduğu için operatöre bağlılık gösterdiği söylenebilir. Şekil 3’de ise RF modelinin %79 olasılıkla sonucu 1 yani mobil operatörden ayrılan olarak tahmin ettiği bir abonenin ELI5, LIME ve SHAP yöntemlerine dayalı lokal açıklanabilirlik sonuçları yer almaktadır.

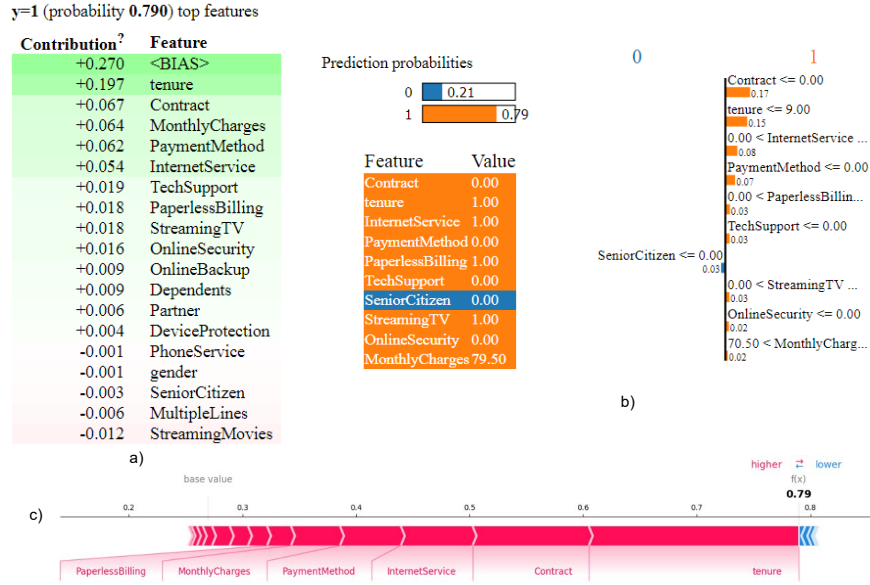
Şekil 3’teki sonuçları incelendiğinde abonenin operatörden ayrılmasında en etkili olan iki özellik sözleşme (contract) süresi, abonenin şirkette kaldığı ay sayısı (tenure) olarak tespit edilmiştir. Diğer özellikler ise sırasıyla aylık ödeme miktarı (MontlyCharges), ödeme yöntemi (PaymentMethod) ve internet servisi (InternetService) olarak devam etmektedir. Detaylı veri analizinde bu abonenin ise aylık/aydan aya (taahütsüz) olarak devam ettiği ve mevcut operatörde 1 aydır bulunduğu görülmüştür. Bu bilgilerden hareketle müşterinin taahhütsüz aylık ödeme planını tercih etmesi, aylık ücretin yüksek olması, ödeme yöntemi ve internet servislerinden kaynaklı memnuniyetsizliği nedeniyle operatör değişikliği yaptığı sonucu çıkarılabilir.



Şekil 2. RF modelinin ELI5, LIME ve SHAP yöntemleriyle churn olmayan/bağlı bir müşteriye ait lokal açıklanabilirlik (a. ELI5 lokal açıklama, b. LIME lokal açıklama, c. SHAP lokal açıklama).

Elde edilen tüm sonuçlar değerlendirildiğinde abonelerin kullandıkları operatörlerde kalma ve operatörü terk etme davranışlarında sözleşme süresi, operatörde kalma süreleri, aylık ücretler, internet servisleri, teknik destek

ve güvenlik gibi kriterler belirleyici etken olarak tespit edilmiştir. Geliştirilen ML algoritmalarının hem global hemde lokal sonuçları bu çıkarımı doğrulamıştır. Bu çalışma sayesinde benzer problemlerin çözümünde sadece ML yaklaşımıyla müşterinin operatörü terk edip etmesi değil, aynı zamanda sonucun neden kaynakladığına ilişkin çıkarımlar elde edilmiştir. Özellikle açıklanabilir yapay zekaya dayalı yöntemler sonuçların yorumlanabilir ve anlaşılabilir olmasına katkı sağlamıştır.



Şekil 3. RF modelinin ELI5, LIME ve SHAP yöntemleriyle churn olan/kaybedilen bir müşteriye ait lokal açıklanabilirlik (a. ELI5 lokal açıklama, b. LIME lokal açıklama, c. SHAP lokal açıklama).

Yukarıda verilen global ve lokal açıklamalar modelin genel ve tek bir örnek bağlamında verdiği sonuçlar göstermektedir. Ancak hizmet sağlayıcılar müşterileri bireysel bazlı olarak elde tutmak için onlara spesifik promosyon ve teklifler sunmak isterler. Bu sebeple her abone özelinde onları yönlendirecek bir takım önerilere ihtiyaç duyarlar. Bu noktada karşıolgusal açıklamalar ilgili ihtiyaca yönelik öneriler sunarlar. Bu bağlamda operatörden ayrılmış (churn=1) bir örnek (Şekil 3'te kullanılan örnek verinin aynısı) için karşı olgusal açıklamalar Tablo 4'te verilmiştir. Tabloda seçilen örnek veriyi RF modeli 1 olarak yani operatörü terk etmiş olarak sınıflandırmıştır. Buna göre karşı olgusal açıklama yöntemiyle bu örneğe en yakın operatörden ayrılmamış dört abone tespit edilmiştir. Bu bilgilere seçilen abonenin operatörü terk etmemesi için karşıolgusal alanda kendinden farklı (kırmızı ile işaretli daire içinde yer alan) özelliklerini tespit etmek ve ayrılan aboneye onlara benzer seçenekler (kampanya, politika, promosyon vb) önermek gerekmektedir. Tablodaki örnek değerlendirildiğinde bu abonenin operatörde kalma süresinin (tenure) artırılmasına yönelik opsiyonlar, internet servisinde kısıtlama, televizyon yayını önerme, sözleşme süresini bir yıla çıkarma ve fatura ödeme yöntemini banka transferi şeklinde yapmasına imkan sağlayan opsiyonlar önerilebilir. Bunlarda uygun olan seçenekler abonenin operatörde kalması sağlayabilir.

Tablo 4. Hizmet sağlayıcı operatörü terk etmiş müşteriye ait karşıolgusal açıklamalar.

Abone Özellikleri	gender	SeniorCitizen	Partner	Dependents	tenure	PhoneService	MultipleLines	InternetService	OnlineSecurity	OnlineBackup	DeviceProtection	TechSupport	StreamingTV	StreamingMovies	Contract	PaperlessBilling	PaymentMethod	MonthlyCharges	Churn
	Seçilen Örnek	1	0	0	0	1	1	0	1	0	0	0	0	1	0	0	1	0	79,5
Counterfactual-1	1	0	0	0	1	1	0	1	0	0	0	0	1	1	1	1	0	80	0
Counterfactual-2	1	0	0	0	11	1	0	2	0	0	0	0	1	0	0	1	0	80	0
Counterfactual-3	1	0	0	0	33	1	0	1	0	0	0	0	1	0	0	1	2	80	0
Counterfactual-4	1	0	0	0	1	1	0	2	0	0	0	0	1	0	1	1	0	80	0

Tablodaki değerlere ilişkin açıklamalar: Tenure: 0-72 ay aralığında değişmektedir. InternetService(0:DSL, 1: FiberOptik, 2: No Service), StreaminTV (0: STV No, 1: STV Yes), Contract (0:Aydan aya 1: bir yıllık 2: iki yıllık) PaymentMethod (0: Elektronik çek, 1:Posta çeki 2: Banka transferi 3: Kredi kartı)

5. SONUÇ

Bu makale mobil telekomünikasyon sektöründeki müşteri kaybı problemine farklı bir perspektif getirmekte ve çözüm sunmaktadır. Bu bağlamda, k-en yakın komşu, destek vektör makinesi, karar ağacı, rastgele orman ve naive bayes gibi makine öğrenimi modelleri geliştirilerek bu modellerin müşteri kaybı tahminlerinde yüksek doğruluk sağladıklarını gösteriyoruz. Bunun yanı sıra, ELI5, LIME, SHAP ve Counterfactual açıklanabilir yapay zeka yöntemleriyle geliştirilen modellerin sonuçlarını hem uzman hem de aboneler anlaşılır ve yorumlanabilir hale getiriyoruz. Bu bulgular, hizmet sağlayıcı operatörlere müşteri davranışlarını daha derinlemesine anlayarak, stratejik kararlar alabilme yeteneği sunmaktadır. Bu çalışmanın en önemli katkılarından biri, hizmet sağlayıcı operatörlere müşteri davranışlarını kapsamlı bir şekilde analiz ederek, müşteri kaybını azaltıcı tedbirler alabilmelerine rehberlik etmektir. Literatürde müşteri kaybını tahmin eden ML modellerinin geliştirilmesini içeren çalışmalar bulunmaktadır. Ancak bu çalışmalar geliştirilen modellerin kararlarına ilişkin açıklamalar ve yorumlanabilirlik sunmamaktadır. Bu nedenle bu çalışmanın literatüre katkısı geliştirilen ML modellerinin açıklanabilir yapay zeka yöntemleriyle desteklenmesidir. Bu sayede, çalışma daha şeffaf ve yorumlanabilir karar süreci sunarak kullanıcıların tahmin modeline ve verilen kararlara olan güvenini artırmaktadır. Diğer taraftan bu çalışmada geleneksel ML algoritmalarına yönelik çalışmalar sunulmuş olması ve model sayısının sınırlı olması bir sınırlılık olarak değerlendirilebilir. Bu bağlamda hem model sayısı hem de daha yüksek doğruluk oranları için derin öğrenme tabanlı modellerin kullanılması gelecek çalışmalar açısından faydalı olacaktır. Sonuç olarak, bu araştırma, telekomünikasyon sektöründeki müşteri kaybı sorununun açılmasında açıklanabilir yapay zeka destekli çözümlerin değerli bir katkı sağlayabileceğini açıkça göstermektedir. Ayrıca bu çalışmada kullanılan model ve yöntemler telekomünikasyon sektöründeki aktörlerin rekabet avantajını elde etmelerine ve sürdürmelerine ilişkin önemli bir gelecek projeksiyonu sunmakta ve müşteri kaybını azaltma potansiyelini ortaya koymaktadır.

Yazar Katkıları

İbrahim Kök: Makalenin kapsamı, yöntemi, araştırmanın uygulanması, sonuçların değerlendirilmesi ve makalenin yazımının tamamına katkı sağlamıştır.

Çıkar Çatışması

Makale yazarları aralarında herhangi bir çıkar çatışması olmadığını beyan ederler.

KAYNAKÇA

- [1] P. Taylor, "Number of smartphone mobile network subscriptions worldwide from 2016 to 2022, with forecasts from 2023 to 2028"i 2023. [Çevrimiçi] Erişim:<https://www.statista.com/statistics/330695/number-of-smartphone-users-worldwide/>
- [2] K. Coussement, S. Lessmann, and G. Verstraeten, "A comparative analysis of data preparation algorithms for customer churn prediction: A case study in the telecommunication industry", *Decis Support Syst*, vol. 95, pp. 27-36, 2017.
- [3] T. Xu, Y. Ma, and K. Kim, "Telecom churn prediction system based on ensemble learning using feature grouping," *Applied Sciences*, vol. 11, no. 11, p. 4742, 2021.
- [4] The European Business Review, "How costly is customer churn in the telecom industry, the European business review", [Çevrimiçi]. Erişim:<https://www.europeanbusinessreview.com/how-costly-is-customer-churn-in-the-telecom-industry/>
- [5] E. Yeboah-Asiamah, B. Narteh, and M.A. Mahmoud, "Preventing Customer Churn in the Mobile Telecommunication Industry: Is Mobile Money Usage the Missing Link?", (in English), *J Afr Bus*, vol. 19, no. 2, pp. 174-194, 2018.
- [6] H. Li et al., "Enhancing Telco Service Quality with Big Data Enabled Churn Analysis: Infrastructure, Model, and Deployment," (in English), *J Comput Sci Tech-Ch*, vol. 30, no. 6, pp. 1201-1214, Nov 2015.
- [7] H. Li, D. L. Yang, L. L. Yang, Y. Lu, and X. L. Lin, "Supervised Massive Data Analysis for Telecommunication Customer Churn Prediction," (in English), *Proceedings of 2016 Ieee International Conferences on Big Data and Cloud Computing (Bdcloud 2016) Social Computing and Networking (Socialcom 2016) Sustainable Computing and Communications (Sustaincom 2016) (Bdcloud-Socialcom-Sustaincom 2016)*, pp. 163-169, 2016.
- [8] M.Z. Kastouni and A. Ait Lahcen, "Big data analytics in telecommunications: Governance, architecture and use cases," *Journal of King Saud University - Computer and Information Sciences*, vol. 34, no. 6, pp. 2758-2770, 2022.
- [9] S. Ali et al., "Explainable Artificial Intelligence (XAI): What we know and what is left to attain Trustworthy Artificial Intelligence," *Inform Fusion*, vol. 99, p. 101805, 2023.
- [10] A. Barredo Arrieta et al., "Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI," *Inform Fusion*, vol. 58, pp. 82-115, 2020/06/01/2020.
- [11] İ. Kök, Y.F. Okay, Ö. Muyanlı, and S. Özdemir, "Explainable Artificial Intelligence (XAI) for Internet of Things: A Survey," *Ieee Internet Things*, vol. 10, no. 16, pp. 14764-14779, 2023.
- [12] P. Kisioglu and Y.I. Topcu, "Applying Bayesian Belief Network approach to customer churn analysis: A case study on the telecom industry of Turkey," (in

- English), *Expert Syst Appl*, vol. 38, no. 6, pp. 7151-7157, Jun 2011.
- [13] N. Lu, H. Lin, J. Lu, and G. Zhang, "A customer churn prediction model in telecom industry using boosting," *Ieee T Ind Inform*, vol. 10, no. 2, pp. 1659-1665, 2012.
- [14] H. Jung, J. Mo, and J. Park, "A Data-Driven Customer Quality of Experience System for a Cellular Network," (in English), *Mob Inf Syst*, vol. 2017, 2017.
- [15] M. Bagri, J.K. Singh, M.K. Abhilash, R.S. Sunitha, and S. Kumar, "Churn Analysis in Telecommunication Industry," (in English), 2018 International Conference on Automation and Computational Engineering (Icace), pp. 126-132, 2018.
- [16] S. Mitrović, B. Baesens, W. Lemahieu, and J. De Weerd, "On the operational efficiency of different feature types for telco Churn prediction," *European Journal of Operational Research*, vol. 267, no. 3, pp. 1141-1155, 2018.
- [17] A. De Caigny, K. Coussement, and K.W. De Bock, "A new hybrid classification algorithm for customer churn prediction based on logistic regression and decision trees," *European Journal of Operational Research*, vol. 269, no. 2, pp. 760-772, 2018.
- [18] T.W. Cenggoro, R.A. Wirastari, E. Rudianto, M.I. Mohadi, D. Ratj, and B. Pardamean, "Deep learning as a vector embedding model for customer churn," *Procedia Computer Science*, vol. 179, pp. 624-631, 2021.
- [19] S.L. Wu, W.C. Yau, T.S. Ong, and S.C. Chong, "Integrated Churn Prediction and Customer Segmentation Framework for Telco Business," (in English), *Ieee Access*, vol. 9, pp. 62118-62136, 2021.
- [20] S.M. Shrestha and A. Shakya, "A Customer Churn Prediction Model using XGBoost for the Telecommunication Industry in Nepal," *Procedia Computer Science*, vol. 215, pp. 652-661, 2022.
- [21] A. Amin, A. Adnan, and S. Anwar, "An adaptive learning approach for customer churn prediction in the telecommunication industry using evolutionary computation and Naïve Bayes," *Applied Soft Computing*, vol. 137, 2023.
- [22] L. Saha, H.K. Tripathy, T. Gaber, H. El-Gohary, and E.-S. M. El-kenawy, "Deep churn prediction method for telecommunication industry," *Sustainability-Basel*, vol. 15, no. 5, p. 4543, 2023.
- [23] B. Prabadevi, R. Shalini, and B.R. Kavitha, "Customer churning analysis using machine learning algorithms," *International Journal of Intelligent Networks*, vol. 4, pp. 145-154, 2023.
- [24] M. Mannan, M.F. Mohiuddin, N. Chowdhury, and P. Sarker, "Customer satisfaction, switching intentions, perceived switching costs, and perceived alternative attractiveness in Bangladesh mobile telecommunications market," (in English), *S Asian J Bus Stud*, vol. 6, no. 2, pp. 142-160, 2017.
- [25] T. Lupo and S.A. Delbari, "A knowledge-based exploratory framework to study quality of Italian mobile telecommunication services," *Telecommun Syst*, vol. 68, no. 1, pp. 129-144, 2018.
- [26] X. Zeng, Y. Wu, and Y. Tang, "Definition of Misclassification Cost & Redistribution Strategy of Telecom Churn Analysis," (in English), 2008 International Conference on Apperceiving Computing and Intelligence Analysis (Icacia 2008), pp. 31-33, 2008.
- [27] J. H. Ahn, S.P. Han, and Y.S. Lee, "Customer churn analysis: Churn determinants and mediation effects of partial defection in the Korean mobile telecommunications service industry," *Telecommun Policy*, vol. 30, no. 10-11, pp. 552-568, 2006.
- [28] IBM. Telco Customer Churn Dataset. [Çevrimiçi]. Erişim: <https://www.kaggle.com/blastchar/telco-customer-churn>
- [29] G. Bonaccorso, *Machine Learning Algorithms: Popular algorithms for data science and machine learning*, Packt Publishing Ltd, 2018.
- [30] M. Mohammed, M. B. Khan, and E. B. M. Bashier, *Machine learning: algorithms and applications*. 1st Edition, Boca Raton, USA, Crc Press, 2016.
- [31] K. P. Murphy, *Machine learning: a probabilistic perspective*. Cambridge, London, UK, MIT press, 2012.
- [32] U. Kamath and J. Liu, *Explainable artificial intelligence: An introduction to interpretable machine learning*, 1st Edition, Switzerland, AG, Springer Cham, 2021.
- [33] S.K. Jagatheesaperumal, Q.-V. Pham, R. Ruby, Z. Yang, C. Xu, and Z. Zhang, "Explainable AI over the Internet of Things (IoT): Overview, state-of-the-art and future directions," *IEEE Open Journal of the Communications Society*, 2022.
- [34] A. Bhattacharya, *Applied Machine Learning Explainability Techniques: Make ML models explainable and trustworthy for practical applications using LIME, SHAP, and more*, Packt Publishing Ltd, 2022.
- [35] S.R. Islam, W. Eberle, S.K. Ghafoor, and M. Ahmed, "Explainable artificial intelligence approaches: A survey," *arXiv preprint arXiv:2101.09429*, 2021.