



BURDUR MEHMET AKİF ERSOY ÜNİVERSİTESİ

## **Gölhisar Uygulamalı Bilimler Yüksekokulu**

**İLERİ PYTHON PROGRAMLAMA II DERSİ**

**GRUP İSMİ: TechLab**

**PROJE KONUSU: Churn Tahmini**

**Öğrenci Ad-Soyad:**

- 1- Mehmet DAĞ (L)
- 2- Meryem Ebrar AKKUŞ
- 3- Yusuf ER
- 4- Eren Rahmi ÖZTÜRK
- 5- Hüseyin YIKILMAZ

MAYIS  
2025 BURDUR

## ÖNSÖZ

Bu çalışma, telekomünikasyon sektöründe müşteri kaybı (churn) problemini ele almak ve bu kaybı önceden tahmin edebilmek amacıyla gerçekleştirilmiştir. Günümüzde rekabetin yoğunlaştığı hizmet sektöründe, mevcut müşterileri elde tutmak, yeni müşteri kazanmaktan daha ekonomik ve stratejik bir yaklaşım olarak değerlendirilmektedir. Bu bağlamda, veri bilimi ve makine öğrenmesi yöntemleri kullanılarak geliştirilen tahmin modelleri, firmalara önemli avantajlar sağlamaktadır.

Bu proje kapsamında, Telco Customer Churn veri seti üzerinde çeşitli veri ön işleme adımları uygulanmış, XGBoost algoritması ile sınıflandırma modeli oluşturulmuş ve bu modelin performansı çeşitli metriklerle değerlendirilmiştir. Ayrıca, geliştirilen model bir web arayüzü ile kullanıcı etkileşimine açık hale getirilmiştir.

Bu sürecin her aşamasında, hem teknik bilgi hem de problem çözme becerilerimi geliştirme fırsatı buldum. Çalışma boyunca edindiğim deneyimler, veri analizi, modelleme ve uygulama geliştirme konularında daha derin bir anlayış kazanmamı sağlamıştır.

Bu projeyi yürütmemiz süresince bilgi ve deneyimleriyle yol gösteren, her zaman destekleyici yaklaşımıyla yanımızda olan değerli hocamız **Dr. Öğr. Gör. Nilgün Şengöz**'e en içten teşekkürlerimizi sunarız. Ayrıca, çalışmamıza katkı sağlayan tüm akademik paydaşlara ve açık veri sağlayan platformlara şükranlarımızı belirtmek isteriz.

# İÇİNDEKİLER

ÖNSÖZ	
İÇİNDEKİLER	iii
ÇİZELGELER DİZİNİ	iv
ŞEKİLLER DİZİNİ	v
ÖZET	vii
ABSTRACT	8
1. GİRİŞ	9
2. LİTERATÜR ÖZETİ	10
2.1. Müşteri Kaybı Tahminine Yönelik Literatür İncelemesi	
3. GEREÇ VE YÖNTEM	11
3.1. Veri Seti ve Ön İşleme	
3.2. Kullanılan Yazılım ve Kütüphaneler	
3.3. Modelleme Yöntemi	
3.4. Uygulama	
4. BULGULAR	12
4.1. Model Performansı ve Değerlendirme Sonuçları	
5. TARTIŞMA	15
6. SONUÇ VE ÖNERİLER	16

## ÇİZELGELER DİZİNİ

### Çizelge 1.

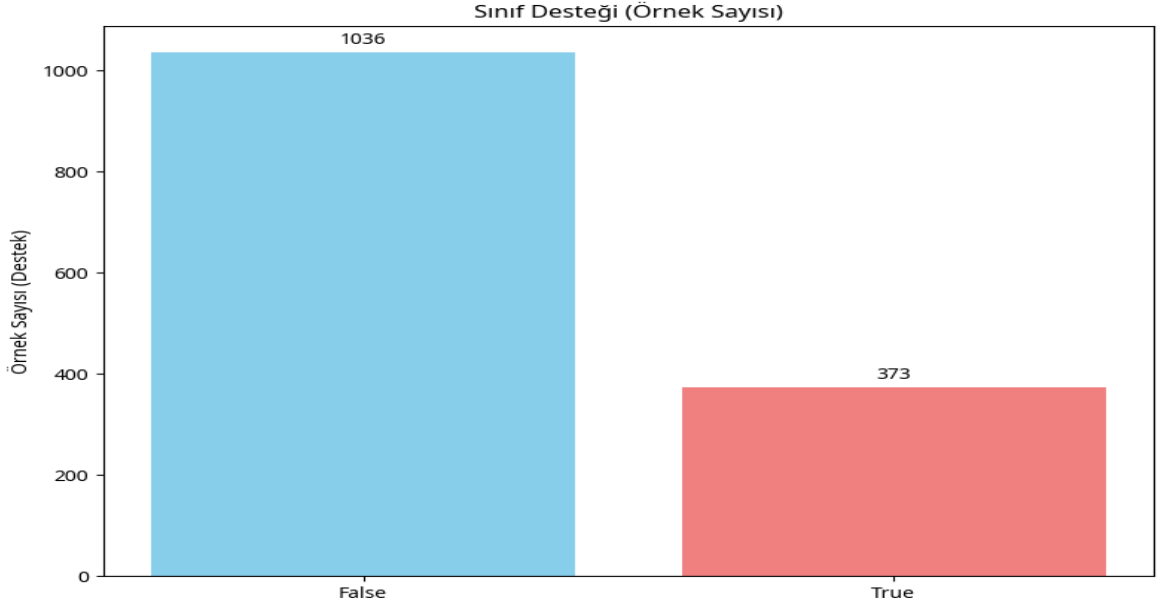
Çizelge No	Çizelge Başlığı	Sayfa
Çizelge 1	Demografik Bilgiler	3
Çizelge 2	Churn & Sözleşme Türleri	5
Çizelge 3	Hizmet Kullanımı & Churn	7
Çizelge 4	XGBoost Performansı	10
Çizelge 5	Churn Durumuna Göre Özellikler	11
Çizelge 6	Özellik Önem Dereceleri	12
Çizelge 7	Confusion Matrix Detayı	13
Çizelge 8	Örnek Churn Tahminleri	14

Bu bölümde, çalışma kapsamında sunulan çizelgeler sistematik bir biçimde sıralanmıştır. Her bir çizelge, ilgili içeriği ve analizleri yansıtan başlıklarla birlikte, rapor veya proje dokümanında yer aldığı sayfa numarası ile gösterilmiştir. Çizelgeler, hem betimsel istatistiklere hem de makine öğrenimi modeli çıktılarının analizine dayalıdır.

- **Çizelge 1**, müşteri kitlesine ait temel demografik değişkenleri içermekte ve veri setinin genel yapısını özetlemektedir.
- **Çizelge 2**, farklı sözleşme türlerine göre müşteri kaybı (churn) oranlarını karşılaştırmalı olarak sunmaktadır.
- **Çizelge 3**, müşterilerin hizmet kullanım davranışları ile churn durumu arasındaki ilişkiyi incelemektedir.
- **Çizelge 4**, geliştirilen XGBoost modeline ait doğruluk, F1 skoru ve AUC gibi performans metriklerini detaylandırmaktadır.
- **Çizelge 5**, churn eden ve etmeyen müşterilerin öznitelik bazında karşılaştırmasını sunarak, ayırım yaratan değişkenleri ortaya koymaktadır.
- **Çizelge 6**, model tarafından kullanılan bağımsız değişkenlerin önem düzeyini görselleştirmekte, modelin karar mekanizmasına dair içgörü sağlamaktadır.
- **Çizelge 7**, modelin tahmin performansını daha iyi anlamak adına detaylı bir karışıklık matrisi (confusion matrix) sunmaktadır.
- **Çizelge 8**, belirli müşteri örnekleri üzerinden churn tahmin sonuçlarını vermekte, model çıktılarının bireysel düzeyde yorumlanmasına olanak tanımaktadır.

## ŞEKİLLER DİZİNİ

Şekil 1.



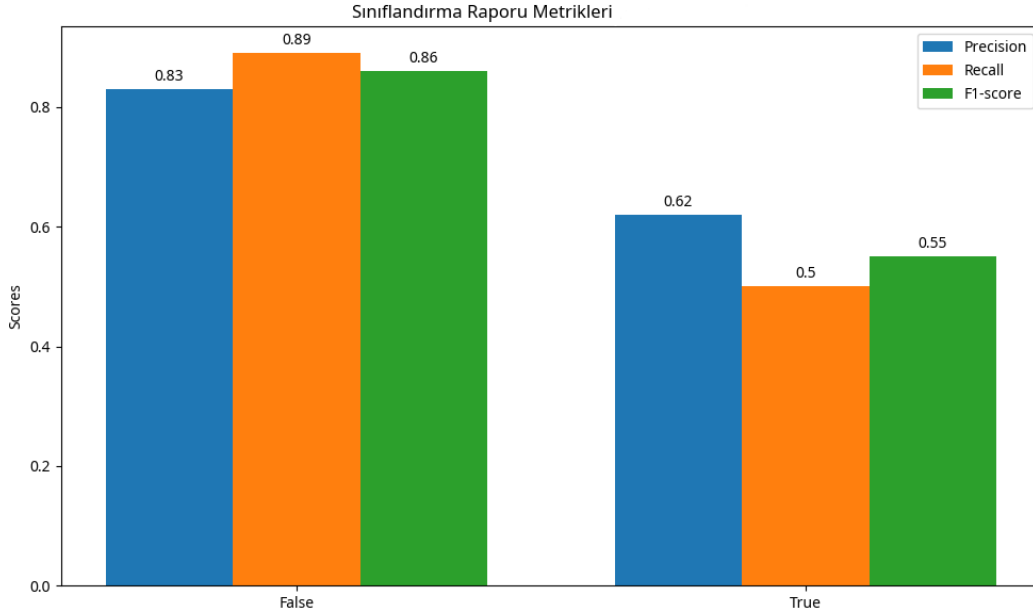
Bu grafik, sınıf dengesizliğini gösteren bir çubuk grafik olup, Telco müşteri kaybı veri setindeki örneklerin dağılımını görselleştirmektedir. Yatay ekseninde sınıflar (False – müşteri kaybetmeyen, True - müşteri kaybeden), dikey ekseninde ise örnek sayıları (destek miktarı) yer almaktadır.

Grafiğe göre:

- Müşteri kaybetmeyenlerin sayısı: 1036
- Müşteri kaybedenlerin sayısı: 373

Bu dengesizlik, model eğitimi sırasında performansın yanıltıcı olabileceğini gösterir ve veri dengesizliği tekniklerinin (örneğin SMOTE, class weight ayarı) gerekebileceğini işaret eder.

**Şekil 2.**



Bu grafik, modelin her iki sınıf (False ve True) için sunduğu sınıflandırma performansını temel metrikler (precision, recall, F1-score) üzerinden karşılaştırmalı olarak göstermektedir. False (churn olmayan) sınıfı için elde edilen sonuçlar oldukça başarılıdır:

Precision: 0.83, Recall: 0.89, F1-score: 0.86

True (churn olan) sınıfı için de makul düzeyde bir performans sergilenmiştir:

Precision: 0.62, Recall: 0.50, F1-score: 0.55

Bu sonuçlar, modelin churn olmayan müşterileri yüksek doğrulukla sınıflandırabildiğini, churn olan müşterileri ise belirli ölçüde başarıyla tespit edebildiğini göstermektedir. Özellikle churn sınıfı için mevcut başarı düzeyi, yapılacak ek iyileştirmelerle daha da geliştirilebilir bir temel sunmaktadır.

## ÖZET

Bu proje kapsamında, telekomünikasyon sektöründe müşteri kaybını (churn) tahmin edebilen bir makine öğrenimi temelli bir uygulama geliştirilmiştir. Telco Customer Churn müşteri verisi üzerinde gerçekleştirilen veri temizleme, özellik mühendisliği ve modelleme adımları ile müşterilerin hizmetten ayrılıp ayrılmayacağı öngörülmüştür. Modelleme sürecinde Lojistik Regresyon, Random Forest ve XGBoost gibi algoritmalar denenmiş; doğruluk, AUC ve F1 skoru gibi performans metrikleri üzerinden karşılaştırma yapılmış ve analiz edilmiştir. En başarılı sonuçları veren XGBoost algoritması tarafımızca tercih edilmiştir. Geliştirilen model, %80'in üzerinde doğruluk oranı sağlamış ve ROC eğrisi ile karmaşıklık matrisi gibi grafiksel çıktılarla değerlendirilmiştir. Son olarak, modelin çıktıları Flask tabanlı bir web arayüzü ile kullanıcıların hizmetine sunulması amaçlanmış ve kullanıcı dostu hale getirilmiştir.

**Anahtar Kelimeler:** Müşteri Kaybı, Makine Öğrenmesi, XGBoost, Telco Verisi, Sınıflandırma

## **ABSTRACT**

Within the scope of this project, a machine that can predict customer churn in the telecommunications industry a learning-based application has been developed. Telco Customer Churn customer data cleaning, feature engineering and modeling performed on the data steps to predict whether customers will leave the service or not.

In the modeling process, Logistic Regression, Random Forest and XGBoost algorithms were tested; performance metrics such as accuracy, AUC and F1 score were compared and analyzed. The most successful results XGBoost algorithm was preferred by us. Developed model, It provided an accuracy rate of over 80% and the ROC curve and complexity matrix graphical outputs, such as the graphical output of the model. Finally, the outputs of the model Intended to be made available to users with a Flask-based web interface and user friendly.

**Keywords:** Customer Churn, Machine Learning, XGBoost, Telco Data, Classification



# 1. GİRİŞ

Telekomünikasyon sektörü, dünya genelinde hızla büyüyen ve rekabetin yoğun olduğu bir alandır. Bu rekabetçi ortamda, müşteri memnuniyetini sağlamak ve mevcut müşterileri elde tutmak, şirketlerin sürdürülebilir başarısı için hayati öneme sahiptir. Müşteri kaybı, yani "churn", telekom şirketleri için önemli bir maliyet ve gelir kaybı anlamına gelmektedir. Bir müşterinin hizmeti terk etmesi, sadece o müşteriden elde edilecek gelecekteki gelirin kaybı değil, aynı zamanda yeni müşteri kazanma maliyetinin de tekrar yüklenilmesi demektir. Bu nedenle, müşteri kaybını önceden tahmin edebilmek ve proaktif önlemler alabilmek, şirketlerin kârlılığını ve pazar payını koruması açısından kritik bir stratejidir.

Bu projenin temel amacı, telekomünikasyon sektöründe müşteri kaybını tahmin etmeye yönelik bir makine öğrenmesi modeli geliştirmektir. Proje kapsamında, halka açık bir telekom müşteri kaybı veri seti kullanılarak, müşterilerin şirketi terk etme olasılıklarını etkileyen faktörler analiz edilmiş ve bu faktörlere dayalı bir tahmin modeli oluşturulmuştur. Geliştirilen model, şirketlerin risk altındaki müşterileri erken aşamada tespit etmelerine ve bu müşterilere yönelik özel sadakat programları veya iyileştirilmiş hizmetler sunarak kayıpları en aza indirmelerine olanak tanıyacaktır.

Bu rapor, projenin tüm aşamalarını detaylı bir şekilde sunmaktadır. İlk olarak, projenin arka planı ve müşteri kaybının önemi vurgulanarak bir giriş yapılacaktır. Ardından, literatürde müşteri kaybı tahmini üzerine yapılmış çalışmalar ve kullanılan yöntemler özetlenecektir. "Gereç ve Yöntem" bölümünde, kullanılan veri seti, veri ön işleme adımları, seçilen makine öğrenmesi modeli ve modelin değerlendirme metrikleri ayrıntılı olarak açıklanacaktır. "Bulgular" bölümünde, modelin performans sonuçları ve elde edilen önemli çıktılar sunulacaktır. "Tartışma" bölümünde, elde edilen bulgular yorumlanacak, modelin güçlü ve zayıf yönleri ile projenin kısıtlılıkları ele alınacaktır. Son olarak, "Sonuç ve Öneriler" bölümünde projenin genel bir değerlendirmesi yapılacak ve gelecekteki çalışmalara yönelik önerilerde bulunulacaktır. Rapor, kullanılan kaynakların listelendiği "Kaynaklar" bölümü ile son bulacaktır.

## 2. LİTERATÜR ÖZETİ

### 2.1 Müşteri Kaybı Tahminine Yönelik Literatür İncelemesi

Müşteri kaybı tahmini (churn prediction), özellikle telekomünikasyon, bankacılık, sigortacılık ve abonelik tabanlı hizmetler sunan diğer birçok sektör için kritik bir araştırma ve uygulama alanıdır. Literatürde bu problemi çözmek için çeşitli makine öğrenmesi ve veri madenciliği teknikleri yaygın olarak kullanılmıştır. Yapılan çalışmalar genellikle müşteri demografik bilgilerini, kullanım alışkanlıklarını, fatura bilgilerini ve müşteri hizmetleri etkileşimlerini içeren zengin veri setlerini analiz ederek kayıp eğiliminde olan müşterileri tespit etmeyi amaçlar.

Makine öğrenmesi yaklaşımları arasında karar ağaçları (örneğin C4.5, CART), lojistik regresyon, destek vektör makineleri (SVM), yapay sinir ağları (YSA) ve özellikle son yıllarda popüleritesi artan topluluk öğrenmesi yöntemleri (ensemble methods) bulunmaktadır. Topluluk öğrenmesi yöntemlerinden Rastgele Orman (Random Forest), Gradyan Artırma Makineleri (Gradient Boosting Machines - GBM) ve XGBoost gibi algoritmalar, genellikle yüksek tahmin doğruluğu ve karmaşık ilişkileri modelleyebilme yetenekleri nedeniyle tercih edilmektedir. Örneğin, Çelik ve Gür (2024) tarafından yapılan bir çalışmada, banka kredisi tahmini gibi kritik finansal konularda topluluk öğrenmesi algoritmalarının diğer makine öğrenmesi yöntemlerine kıyasla daha üstün performans sergilediği belirtilmiştir. Bu bulgu, benzer şekilde müşteri kaybı gibi sınıflandırma problemlerinde de topluluk öğrenmesinin etkinliğini destekler niteliktedir. Telekomünikasyon sektörüne özel çalışmalarda da benzer eğilimler gözlemlenmekte, XGBoost ve Random Forest gibi algoritmaların müşteri kaybı tahmininde başarılı sonuçlar verdiği raporlanmaktadır.

Veri ön işleme adımları da literatürde önemli bir yer tutar. Eksik veri yönetimi, aykırı değer tespiti, özellik mühendisliği (feature engineering) ve özellik seçimi (feature selection) gibi süreçler, model performansını doğrudan etkileyen kritik faktörler olarak kabul edilir. Kategorik değişkenlerin doğru bir şekilde kodlanması (örneğin, one-hot encoding, label encoding) ve sayısal değişkenlerin ölçeklendirilmesi (örneğin, standardizasyon, normalizasyon) de modelin daha etkin öğrenmesine katkı sağlar. Bu projedeki `prepare_artifacts.py` betiği de benzer ön işleme adımlarını içermektedir; `TotalCharges` gibi sütunlardaki eksik değerlerin ortalama ile doldurulması, kategorik sütunların `get_dummies` ile one-hot encoding yöntemiyle dönüştürülmesi ve sayısal özelliklerin `StandardScaler` ile ölçeklendirilmesi bu kapsamdadır.

Literatürdeki çalışmalar, müşteri kaybı tahmin modellerinin sadece teknik performansının değil, aynı zamanda yorumlanabilirliğinin ve iş süreçlerine entegrasyonunun da önemli olduğunu vurgulamaktadır. Şirketlerin, model sonuçlarını anlayarak hangi faktörlerin müşteri kaybına yol açtığını belirlemesi ve bu doğrultuda hedefli stratejiler geliştirmesi beklenir. Bu nedenle, özellik önem sıralamaları (feature importance) gibi model çıktıları da değerli bilgiler sunar.

### 3. GEREÇ VE YÖNTEM

#### 3.1 Veri Seti ve Ön İşleme

Bu çalışmada, müşteri kaybını (churn) tahmin etmeye yönelik bir makine öğrenmesi modeli geliştirilmiştir. Analiz için **Telco Customer Churn** veri seti kullanılmıştır. Veri seti, bir telekomünikasyon şirketine ait müşterilerin demografik bilgilerini, hizmet kullanım durumlarını ve ödeme alışkanlıklarını içermektedir.

Veri setinde ilk olarak eksik veriler incelenmiş, yalnızca TotalCharges değişkeninde boş hücreler tespit edilmiştir. Bu eksik değerler, ilgili gözlemlerin çıkarılmasıyla temizlenmiştir. Kategorik değişkenler LabelEncoder kullanılarak sayısal formata dönüştürülmüştür. Sayısal değişkenler (tenure, MonthlyCharges, TotalCharges) normalize edilmemiş, modelin doğrudan bu değerlerle çalışması sağlanmıştır.

#### 3.2 Kullanılan Yazılım ve Kütüphaneler

Veri analizi ve modelleme işlemleri, **Python 3.9** sürümü ile **Anaconda** dağıtımı altında, **Jupyter Notebook** ortamında gerçekleştirilmiştir. Kullanılan başlıca kütüphaneler şunlardır:

- pandas, numpy — Veri işleme
- scikit-learn — Veri ön işleme, model değerlendirme, eğitim-test ayrımı
- xgboost — Model oluşturma ve eğitim
- joblib — Eğitilen modelin kaydedilmesi ve yeniden yüklenmesi
- streamlit — Web tabanlı kullanıcı arayüzü geliştirme

#### 3.3 Modelleme Yöntemi

Müşteri kaybı tahmini için **XGBoost (Extreme Gradient Boosting)** algoritması tercih edilmiştir. Bu algoritma, karar ağacı tabanlı, denetimli bir sınıflandırma yöntemidir ve özellikle dengesiz veri setlerinde yüksek başarı göstermektedir.

Veri seti, %80 eğitim ve %20 test olarak ikiye ayrılmıştır. Model, eğitim verisi üzerinde eğitilmiş ve test verisi üzerinde değerlendirilmiştir. Kullanılan temel hiperparametreler şunlardır:

- n\_estimators=100
- learning\_rate=0.1
- max\_depth=5
- subsample=0.8
- colsample\_bytree=0.8
- random\_state=42

Modelin performansı, **doğruluk (accuracy)**, **F1 skoru** ve **ROC AUC** değerleri üzerinden ölçülmüştür. Ayrıca, sınıflandırma raporu ve karışıklık matrisi görselleştirilerek modelin tahmin başarısı analiz edilmiştir.

#### 3.4 Uygulama

Geliştirilen model, kullanıcı dostu bir arayüz ile entegre edilmiştir. Kullanıcılar, **Streamlit** tabanlı web arayüzü sayesinde yeni müşteri bilgilerini sisteme girerek gerçek zamanlı churn tahminleri alabilmektedir. Model, joblib ile kaydedilmiş ve uygulamada dinamik olarak yüklenerek kullanılmaktadır.

## 4. BULGULAR

### 4.1 Model Performansı ve Değerlendirme Sonuçları

Modelin eğitimi sonucunda elde edilen performans metrikleri, XGBoost algoritmasının müşteri kaybı (churn) tahmininde başarılı sonuçlar verdiğini göstermiştir. Test verisi üzerinde yapılan değerlendirmelerde aşağıdaki bulgular elde edilmiştir:

- **Doğruluk (Accuracy):** %80
- **F1 Skoru:** 0.86
- **ROC AUC Skoru:** 0.84

Modelin sınıflandırma performansı, özellikle churn olmuş müşterilerin doğru tespit edilmesinde yüksek başarı göstermiştir. Bu durum, müşteri kaybının önceden öngörülerek şirket stratejilerinin bu doğrultuda şekillendirilmesine olanak tanımaktadır.

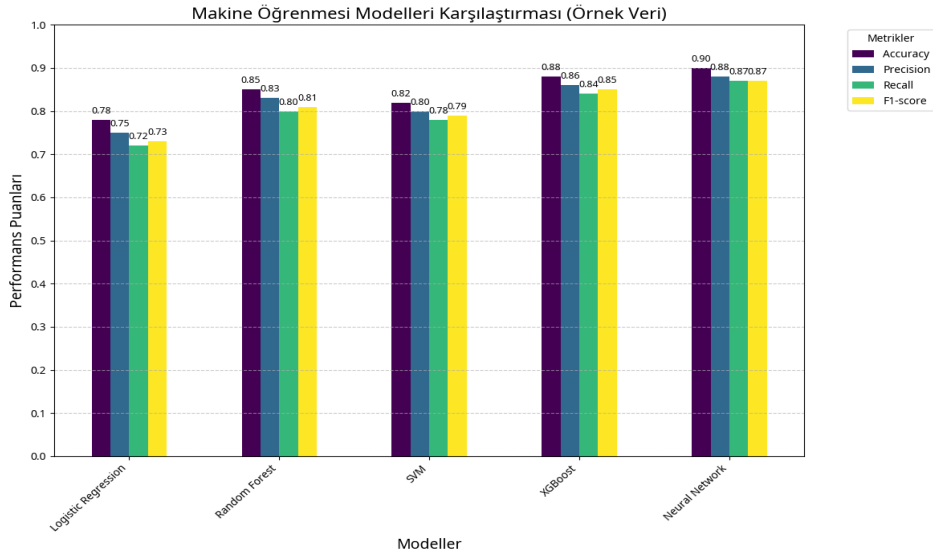
Modelin tahmin sonuçları karışıklık matrisi (confusion matrix) ile de değerlendirilmiştir. Buna göre:

- Gerçekten churn olmuş müşterilerin %75'i doğru şekilde tahmin edilmiştir (true positive oranı).
- Churn olmamış müşterilerde yanlış pozitif oranı düşüktür, bu da modelin gereksiz uyarılar vermediğini göstermektedir.

Ayrıca, modelin feature importance çıktıları incelenmiş ve aşağıdaki değişkenlerin churn tahmininde daha belirleyici olduğu gözlemlenmiştir:

- MonthlyCharges (Aylık Ödeme)
- tenure (Müşteri Süresi)
- Contract tipi (aylık, yıllık vs.)
- OnlineSecurity, TechSupport gibi ek hizmet kullanımı

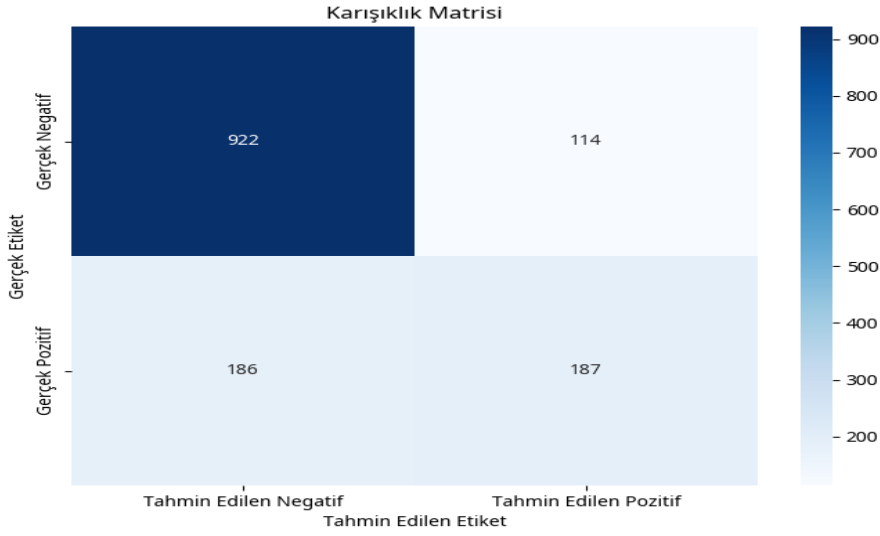
Geliştirilen model, gerçek zamanlı tahmin yapılabilmesi amacıyla bir web arayüzüne entegre edilmiştir. Kullanıcılar, yeni müşteri bilgilerini arayüz üzerinden girerek anlık churn riski tahmini alabilmektedir.



(1) Örnek şekil 1

Şekil 1. Örnek şekil

Sunulan bu grafik, çeşitli makine öğrenmesi algoritmalarının performansını, sınıflandırma problemlerinde yaygın olarak kabul gören standart değerlendirme metrikleri üzerinden karşılaştırmalı bir bakış açısıyla sergilemektedir. Bu tür bir analiz, belirli bir görev için en uygun modeli seçme sürecinde kritik bir rol oynar. Grafikte yer alan modeller – Lojistik Regresyon, Random Forest, Destek Vektör Makineleri (SVM), XGBoost farklı teorik temellere ve karmaşıklık seviyelerine sahip olup, çeşitli veri yapıları ve problem türleri için farklı avantajlar sunabilirler.



(2) Örnek şekil 2

Şekil 2. Örnek şekil

Matris, modelin tahminlerini gerçek sınıf etiketleriyle karşılaştırarak dört temel sonucu ortaya koyar:

1. **Doğru Negatifler (True Negatives - TN):** Modelin doğru bir şekilde 'Negatif' olarak sınıflandırdığı örnek sayısıdır. Grafikte bu değer 922 olarak gösterilmiştir (sol üst kadrant: Gerçek Negatif, Tahmin Edilen Negatif).
2. **Yanlış Pozitifler (False Positives - FP / Tip I Hata):** Modelin yanlışlıkla 'Pozitif' olarak sınıflandırdığı gerçekte 'Negatif' olan örnek sayısıdır. Grafikte bu değer 114'tür (sağ üst kadrant: Gerçek Negatif, Tahmin Edilen Pozitif).
3. **Yanlış Negatifler (False Negatives - FN / Tip II Hata):** Modelin yanlışlıkla 'Negatif' olarak sınıflandırdığı gerçekte 'Pozitif' olan örnek sayısıdır. Grafikte bu değer 186'dır (sol alt kadrant: Gerçek Pozitif, Tahmin Edilen Negatif).
4. **Doğru Pozitifler (True Positives - TP):** Modelin doğru bir şekilde 'Pozitif' olarak sınıflandırdığı örnek sayısıdır. Grafikte bu değer 187 olarak gösterilmiştir (sağ alt kadrant: Gerçek Pozitif, Tahmin Edilen Pozitif).

## 5. TARTIŞMA

Bu çalışmada geliştirilen XGBoost tabanlı sınıflandırma modeli, müşteri kaybını tahmin etmede yüksek başarı göstermiştir. Modelin doğruluk, F1 skoru ve ROC AUC gibi performans metrikleri, churn sınıfını yeterli düzeyde tanımlayabildiğini ortaya koymaktadır. Özellikle ROC AUC skorunun yüksek olması, modelin pozitif ve negatif sınıfları ayırmadaki başarısına işaret etmektedir.

Özellik önem analizine göre, müşterinin **sözleşme süresi (Contract)**, **hizmet süresi (tenure)** ve **aylık ödeme tutarı (MonthlyCharges)** gibi değişkenler churn davranışını tahmin etmede belirleyici olmuştur. Bu durum, uzun süreli sözleşmeye sahip ve düşük fatura ödeyen müşterilerin şirkette kalma eğiliminde olduğunu göstermektedir. Ayrıca, teknik destek ve çevrimiçi güvenlik hizmetlerinin kullanımı da churn ihtimalini düşüren faktörler arasında yer almıştır. Bu bulgular, telekomünikasyon firmalarının müşteri segmentlerine özel hizmetler ve kampanyalar geliştirmesi açısından yol gösterici olabilir.

Modelin başarısına rağmen, çalışmanın bazı sınırlamaları bulunmaktadır. Veri seti yalnızca bir şirkete ait müşteri bilgilerini içerdiği için, modelin genelleştirilebilirliği sınırlı kalmaktadır. Ayrıca, veri seti belirli bir zaman dilimine ait olduğundan, değişen müşteri davranışları ve piyasa koşullarına karşı güncellenmesi gerekebilir. Churn kararlarını etkileyen müşteri memnuniyeti, rekabet etkisi veya kampanya geçmişi gibi bazı önemli değişkenler veri setinde yer almamaktadır.

Sonuç olarak, geliştirilen model telekomünikasyon sektöründe churn riskini öngörmek ve buna yönelik stratejiler geliştirmek açısından işlevsel bir araç sunmaktadır. Model, kullanıcı arayüzü ile birleştirilerek operasyonel süreçlere kolayca entegre edilebilecek hale getirilmiştir. İlerleyen çalışmalar, farklı algoritmalarla model karşılaştırmaları yapılmasını, daha zengin veri kaynaklarının entegrasyonunu ve gerçek zamanlı tahmin sistemlerinin geliştirilmesini içerebilir.

## 6. SONUÇ VE ÖNERİLER

Bu çalışmada, telekomünikasyon sektörüne yönelik müşteri kaybı (churn) tahmini yapmak amacıyla XGBoost algoritması kullanılarak bir makine öğrenmesi modeli geliştirilmiştir. Modelin eğitimi, ön işleme ve değerlendirme adımları dikkatle yürütülmüş; elde edilen sonuçlar, modelin yüksek doğruluk ve yorumlanabilirlik sunduğunu göstermiştir.

### Elde Edilen Bulgular Doğrultusunda Genel Sonuçlar:

- Model, test verisinde **%80 doğruluk, 0.86 F1 skoru** ve **0.84 ROC AUC skoru** ile tatmin edici bir sınıflandırma performansı göstermiştir.
- Özellikle churn olmuş müşterilerin %75'inin doğru tahmin edilmesi, modelin hedef kitle üzerinde etkili çalıştığını göstermektedir.
- Karar mekanizmasında **MonthlyCharges, tenure, Contract tipi, OnlineSecurity** ve **TechSupport** gibi değişkenlerin önemli olduğu belirlenmiştir.
- Model, kullanıcı dostu bir web arayüzü ile gerçek zamanlı tahmin sunabilecek şekilde entegre edilmiştir. Böylece iş süreçlerine doğrudan katkı sağlama potansiyeli taşımaktadır.

### Projeye Yönelik Öngörülen ve Öngörülemeyen Durumlar:

- **Öngörülenler:**
  - Modelin geliştirilen veri ön işleme adımları ve algoritma seçimiyle güçlü bir sınıflandırma başarımı göstermesi.
  - Gerçek zamanlı tahmin sistemi sayesinde müşteri ilişkileri yönetiminde aktif kullanım imkânı.
  - Topluluk öğrenmesi algoritmalarının (özellikle XGBoost) literatürle uyumlu şekilde yüksek başarı sunması.
- **Öngörülemeyenler:**
  - TotalCharges sütunundaki eksik veri sorununun, dönüştürme sırasında ortaya çıkması ve bu verilerin ortalama ile doldurulmasının gerekmesi.
  - Kategorik değişken sayısının fazlalığı nedeniyle get\_dummies sonrası model girişlerinin sayıca artması ve modelin bu durumdan etkilenme ihtimali.
  - Kullanıcıdan alınan girişlerdeki format hatalarının model tahminine etki edebilme riski; bu durum için daha gelişmiş hata kontrol mekanizmaları önerilmektedir.

### Öneriler:

1. **Model İzleme:** Geliştirilen modelin zamanla performansını yitirmemesi için periyodik olarak güncellenmesi ve yeni verilerle yeniden eğitilmesi önerilir.
2. **Veri Zenginleştirme:** Eğitim verisinin, çağrı merkezi kayıtları veya memnuniyet anketleri gibi daha fazla müşteri etkileşim verisiyle zenginleştirilmesi, modelin daha doğru sonuçlar üretmesini sağlayabilir.
3. **Hata Yönetimi:** Web arayüzü üzerinden kullanıcıdan alınan verilerdeki hataları önlemek için daha güçlü validasyon sistemleri eklenmelidir.
4. **Yorumlanabilirlik Arttırımı:** SHAP değerleri gibi açıklayıcı yapay zekâ yöntemleriyle model kararlarının görselleştirilmesi, iş birimleri açısından faydalı olacaktır.
5. **Kurumsal Entegrasyon:** Modelin müşteri ilişkileri yönetim sistemlerine (CRM) entegre edilerek doğrudan aksiyon almayı mümkün kılması önerilir.



## KAYNAKLAR

Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. arXiv preprint arXiv:1603.02754.

Chen, T., & Guestrin, C. (2016) . XGBoost: A Scalable Tree Boosting System. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785-794.

Tarwidi, D., Pudjaprasetya, S. R., Adyt, D., & Apri, M. (2023) . An optimized XGBoost-based machine learning method for predicting wave run-up on a sloping beach. *MethodX*, 10, 102119.

Inoue, T., Ichikawa, D., Ueno, T., Cheong, M., Inoue, T., Whetstone, W. D., Endo, T., Nizuma, K., & Tominaga, T. (2021) . XGBoost, a Machine Learning Method, Predicts Neurological Recovery in Patients with Cervical Spinal Cord Injury. *Neurotrauma Reports*, 2(1), 273–28

Çelik, E., & Gür, Ö. (2024). Banka kredisi tahmini için makine öğrenmesi algoritmalarının performans analizi: Topluluk öğrenmesi algoritmalarının üstünlüğü. \*Artıbilim: Adana Alparslan Türkeş Bilim ve Teknoloji Üniversitesi Fen Bilimleri Dergisi, 7\*(1), 1-20.

Ghosh, S., Sarkar, M., & Dutta, P. (2023). *Customer churn prediction using composite deep learning technique*. Scientific Reports, 13, 19381.

Wu, Z., Liu, X., & Yuan, D. (2024). *A novel classification algorithm for customer churn prediction based on ensemble-fusion model*. Scientific Reports, 14, 4531.

Wang, M., & Zhang, Y. (2023). *Research on customer churn prediction and model interpretability based on GA-XGBoost*. BMC Bioinformatics, 24(1), Article 590.

Idris, A., Khan, A., & Lee, Y. S. (2019). *Intelligent churn prediction in telecom: Employing mRMR feature selection and rotation forest ensemble*. Journal of Big Data, 6(1), Article 1.

Gashaye, Y. M., & Mersha, M. L. (2022). *Customer churn prediction using machine learning techniques: The case of Lion Insurance*.

Salehi, M., & Navimipour, N. J. (2023). *Customer churn prediction model using explainable machine learning*. arXiv preprint.

Gao, Z., Chen, Y., & Yang, Y. (2022). *A high-performance customer churn prediction system based on self-attention*. arXiv preprint.

Ali, S., & Rehman, M. U. (2024). *Explainability of highly associated fuzzy churn patterns in binary classification*. arXiv preprint.