

Açıklanabilir Yapay Zekâ Kullanarak Satın Alma Davranışlarının Açıklanması

Explaining Purchasing Behaviors via XAI

Necati DOĞRUL

Akdeniz Üniversitesi, Bilgisayar Mühendisliği, necatidogrul7@gmail.com

Özet

Bu makalede, müşteri satın alma davranışlarını anlamak ve tahmin etmek için makine öğrenmesi ve açıklanabilir yapay zekâ (AYZ) teknikleri kullanılmaktadır. Müşteri davranışlarının anlaşılması ve satın alma örüntülerinin bulunması, müşterilerin hangi sıklıkla ne ürünlere ilgi duyduğunu, alınan ürünün neden alındığını açıklayan bilgiler sunabilir. AYZ, modellerin karar verme sürecini daha açık ve anlaşılabilir hale getirerek bu amaca ulaşmayı sağlar. Makalede RandomForestClassifier, XGBoost ve CatBoost gibi makine öğrenmesi algoritmalarından yararlanılmıştır. Bu algoritmalar arasında en yüksek başarı oranı RandomForestClassifier'a aittir, be doğruluk (accuracy) oranı %99,21'dir. CatBoost ise en yüksek 2.doğruluk oranına sahiptir, bu da %99,08'dir. XGBoost'un ise doğruluk oranı %99,04'tür. Müşterinin satın alma alışkanlıkları için 216 farklı kombinasyonda çıktı oluşturulmuştur. Müşterinin satın alma alışkanlıklarına göre potansiyel bir çıktı üretilmektedir. Bu makale sayesinde işletmeler satın alma örüntülerini kullanarak kampanyalarını yönetebilir, müşterilerin günlere göre ne alacaklarını tahmin edebilirler ve fiyat stratejilerini geliştirebilirler. İndirimlerin müşteri satın alma davranışları üzerindeki etkisini daha iyi anlayarak, pazarlama stratejilerini optimize edebilirler.

Anahtar kelimeler: Açıklanabilir yapay zekâ, XAI, Müşteri satın alma davranışları, Makine öğrenmesi, satın alma örüntüleri

Abstract

This article uses machine learning and explainable artificial intelligence (XAI) techniques to understand and predict customer purchasing behavior. Understanding customer behavior and finding purchasing patterns can provide information explaining how often customers are interested in what products and why the product is purchased. XAI achieves this goal by making the decision-making process of models more clear and understandable. Machine learning algorithms such as RandomForestClassifier, XGBoost and CatBoost were used in the article. The highest success rate among these algorithms belongs to RandomForestClassifier, and its accuracy rate is 99.21%. CatBoost has the 2nd highest accuracy rate, which is 99.08%. XGBoost's accuracy rate is 99.04%. Outputs were created in 216 different combinations for the customer's purchasing patterns. A potential output is produced according to the customer's purchasing patterns. Thanks to this article, businesses can manage their campaigns using purchasing patterns, predict what customers will buy by day, and

develop their pricing strategies. By better understanding the impact of discounts on customer purchasing behavior, they can optimize their marketing strategies.

Keywords: Explainable artificial intelligence, XAI, Customer purchasing behavior, Machine learning, purchasing patterns

1 GİRİŞ

Yapay zekâ ve makine öğrenmesi, teknolojinin hızlı gelişimiyle birlikte günümüz dünyasının birçok alanında kullanılmaktadır. Makine öğrenmesi, algoritmaların önce veriye bakarak öğrenmesi ve daha sonra veriden bağımsızca tahmin yapması üzerine kurulmuştur. Bu modellerin performansı dışında şeffaf ve adaletli olması da önemlidir [1].

Yapay zekâ modelleri birçok parametre içerebilir [2]. Bu parametrelerin hepsi insanlar tarafından anlaşılabilir ve bu yüzden, yapay zekalar insanlar tarafından "kara kutu" olarak görülebilir. Yani, bu sistemlerin girdi ve çıktıları arasında oluşan süreç insanlar tarafından anlaşılabilir. Açıklanabilir Yapay Zekâ (AYZ) sayesinde bu sorun çözülebilir ve algoritmaların karar verme sürecinin daha açık ve anlaşılabilir olması sağlanabilir [3]. Bu açıklanabilirlik standartlarının nasıl oluşturulacağı konusunda ortak zeminin kurulmasının önündeki zorluk, açıklanabilir yapay zekâ kavramları arasında farklılıkların bilinmemesinden kaynaklanmaktadır. Bu farklılıkları ve benzerlikleri açıklığa kavuşturmak amacıyla, açıklanabilirlik terminolojisinde sıkça kullanılan kavramlardan bazıları şunlardır: [3]

- Aktarılabirlik: Açıklanabilir yöntemin, önceki bilgileri alışılmadık durumlara aktarabilme kapasitesi
- Anlaşılabilirlik: Kullanılan dilin kalitesi
- Bütünlük: Açıklamaların yeterlilik düzeyi
- Doğruluk: Açıklanabilirlik yönteminin gerçekten alakalı özellikleri seçme kapasitesi
- Geridönenebilirlik: Kötü bir seçimden sonra sistemin tekrar orijinal haline getirilebilmesi
- Nedensellik: Girdi ve çıktı arasındaki ilişkinin netleştirilmesi
- Tutarlılık: Benzer girdiler için benzer açıklamaların sağlanması
- Sadelik: Modelin tahmininin açıklanması için yalnızca gerekli ve nedenlerin seçilmesi

AYZ'nin amacı, model kararlarını anlamlı ve anlaşılır cümleler veya grafikler şeklinde sunmaktır. Bu yüzden, modelin işleyişi şeffaf ve anlaşılabilir şekilde gösterilmelidir [4].

AYZ'nin önemi, üzerinde çalışılan konuya ve sonuca göre değişebilir. Örneğin, bir fotoğrafın, bir uçak mı yoksa bir kuş mu olduğunu belirlemek düşük açıklanabilirlik gerektirir ve bu tanımda modelin iç işleyişini anlamak, doğru tahmin yapılmasına göre daha az öneme sahiptir. Ancak, tıbbi teşhis için oluşturulan modellerde yani insan sağlığının ön planda olduğu uygulamalarda, yüksek düzeyde açıklanabilirlik ihtiyacı vardır [5]. Örneğin, bir model bir hastanın kanser olup olmadığını tahmin ediyorsa, bu tahminin neden, nasıl ve hangi özelliklerin kullanılarak yapıldığının bilinmesi kritik bir öneme sahip olabilir ve hata yapılırsa geri dönülemeyecek sonuçlar doğurabilir, bu da risk içerir [6]. Bu risklerin gerçekleştiği birkaç örnek durum Guidotti ve arkadaşları tarafından şu şekilde özetlenmiştir: [7]

- Sorumluluk ve Etik: Modelin yanlış karar verdiği durumların açıklanamaması, etik sorunlara yol açabilir.
- Kararların Anlaşılması: Modellerin iç işleyişi karışık olduğundan dolayı kararları anlamak zor olabilir.
- Önyargı: Modelin eğitildiği verilerdeki önyargılar, adalet ve eşitlik problemlerine yol açabilir.

Ayrıca, AYZ sayesinde modeli kullanan kişilerin modeli anlaması sağlanır ve bu sayede modelin güvenilirliği artabilir. Bu sayede, yapay zekanın sorumlu ve etik şekilde kullanımını desteklenir [8]. Ayrıca, modelin iç işleyişini anlaşılması ve modelin sonuçlarına güvenilmesi sağlanır [9].

AYZ, modellerin hatalı sonuç verme ihtimalini azaltabilir [7]. AYZ ile, modelin hangi özellikleri kullandığı açıklanabilir ve bu da modelin hatalı tahminler yapmasına yol açabilecek hataların belirlenmesinde yardımcı olabilir. Aynı zamanda AYZ, modelin performansının ve o performans üzerinde etkisi olan özelliklerin neler olduğunun anlaşılmasını sağlar. Bu da o modeli iyileştirme şansı yaratır ve model hata yaptığı zaman, bu hatanın bulunmasına yardımcı olabilir. Bu sayede modelin geliştirilmesi için gerekli adımlar atılabilir. AYZ için birçok teknik vardır. Bu tekniklerden bazıları şunlardır;

- LIME (Local Interpretable Model Explanations) bir makine öğrenmesi modelinin karmaşık tahminlerini yerel olarak açıklamak için bazı modeller kullanır. Bir veri kümesi etrafında birkaç hatalı veri yaratır ve bu verilere göre basit bir model eğitir. Bu basit model, karmaşık modelin kararlarının anlaşılmasını kolaylaştırır [10].
- SHAP (Shapley Additive Explanations) değerleri, bir oyun teorisi kavramı olan Shapley değerlerine dayanır ve her özelliğin modelin çıktısına katkısını açıklar, her özelliğin katkısını tahmin eder ve modelin kararlarını şeffaflaştırır [11].
- Karar Ağaçları, kendiliğinden açıklayıcı bir yapıdadır çünkü veriyi bazı kurallar kullanarak bölümlere ayırır. Karar ağaçları, modelin nasıl kararlar verdiğini ve model özelliklerinin hangisinin daha önemli olduğunu görselleştirir [12].

Yapay zekâ sistemleri, KOBİ (Küçük veya Orta Büyüklükteki İşletmeler) için de kullanılabilir. Sektördeki dijital dönüşüm ve rekabet satın alma davranışlarını öngörmek için daha etkili yöntemlere olan talebi arttırmıştır. Müşterilerin satın alma alışkanlıklarını anlamak ve öngörmek, işletmelerin pazarlama stratejilerini iyileştirmesine olanak sağlar ve bu sayede müşteri memnuniyeti artırılabilir [13,14].

Örneğin Petter Eilif de Lange ve arkadaşlarının makalesinde AYZ, kredi değerlendirme modellerinin tahminlerini anlaşılır bir şekilde göstermek için kullanılmıştır. Bu sayede modelin hangi değişkenlerin müşterilerin tahminlerini nasıl etkilediği daha iyi anlaşılır. Ayrıca, LightGBM modelinin bankanın halihazırda bulunan Kredi Skorlama Modeline göre daha iyi performans gösterdiği belirtilmektedir. Ve bu sayede tahminlerin doğruluğu artırılabilir. AYZ ile elde edilen açıklamalar sayesinde de müşteri davranışlarının daha iyi anlaşılması sağlanır. Bu sayede bankalar veya AYZ kullanan herhangi bir işletme müşteriye özgü stratejiler geliştirebilir ve müşteri ilişkilerini üst seviyeye çıkarır [15].

Felix Haag ve arkadaşları tarafından yazılan bir makalede ise, enerji işletmeleri (perakende) alanında AYZ kullanarak çapraz satışın faydaları incelenmiştir. AYZ, enerji perakendecilerinin müşteri davranışlarını daha iyi tahmin etmelerine ve anlamalarını sağlar. Bu sayede her müşteriye özel, ayrı teklifler sunarak aradaki ilişkinin güçlenmesi sağlanır. Ayrıca AYZ, müşteri tercihlerinin anlaşılmasını sağlar ve gelecekteki satışların daha iyi tahmin edilmesini sağlayabilir. Aynı zamanda AYZ, müşteri profillerini analiz eder ve çapraz satış yapılacak ürün ve hizmetlerin belirlenmesini sağlar. Bu da müşterilere daha fazla seçenek sunar. Enerji perakendeciliğinde AYZ sistemlerini kullanmak, rakiplerden ayrılma ve müşteri sadakatini artırma fırsatı sunabilir. Müşterilere daha iyi hizmet ve öneriler sunulabilir ve bu da müşteri sadakatini artırarak rekabet avantajı sağlayabilir [16].

Saavi Stubseid ve arkadaşı tarafından yazılan makalede ise makine öğrenmesi tekniklerinin tüketici satın alma davranışlarını tahmin etmek için nasıl kullanılabileceği incelenmektedir ve makine öğrenmesinin bu alandaki artan önemi ve potansiyeli vurgulanmaktadır. Ayrıca, makine öğrenmesinin tüketici satın alma davranışlarını tahmin etmek için kullanıldığı çeşitli çalışmalara göz atılmaktadır. Bu çalışmalarda, makine öğrenmesinin demografik bilgiler, satın alma geçmişi ve ürün özellikleri gibi değişkenleri kullanarak tüketici satın alma kararlarını tahmin etmek için etkili bir araç olduğu gözlemlenmiştir [17].

Makale, makine öğrenmesinin tüketici satın alma davranışlarını tahmin etmek için kullanılmasının hem teorik hem de pratik açıdan önemli olduğunu savunmaktadır. Teorik olarak, makine öğrenmesi, tüketici davranışının daha iyi

anlaşılmasına ve tüketici odaklı pazarlama stratejilerinin geliştirilmesine yardımcı olabilir. Pratik olarak, makine öğrenmesi, işletmelerin müşterilerin satın alma davranışlarını tahmin ederek daha iyi ürün ve hizmetler sunmasına ve satışlarını arttırmasına yardımcı olabilir. Makalede makine öğrenmesi modeli olarak sınıflandırma algoritmalarından Naive Bayes ve Random Forest kullanılmış. Ve bu modellerin başarısını ölçmek için Accuracy, AUC ve F1-score değerlendirme araçlarından yararlanılmış. Tüm değerler Tablo 1’de gösterilmiştir.

Tablo 1. Stubseid S’nin makalesinin değerlendirme sonuçları

Measure	Naive Bayes	Random Forest
Accuracy	0.66	0.72
AUC	0.71	0.79
F1-score	0.66	0.72

Makalenin sonunda, yazarlar makine öğrenmesinin tüketici satın alma davranışlarını tahmin etmek için kullanılmasının önündeki bazı zorluklara dikkat çekiyor. Bu zorluklar, veri eksikliği, veri kirliliği ve model karmaşıklığıdır.

Choudhury’nin yazdığı makalede ise makine öğrenmesi sayesinde potansiyel müşterilerini belirlemesine, onları daha iyi anlamasına ve satış miktarını arttırma konusunda yardımcı olmuştur. Ve bu modeli diğer işletmelerin de uygulayabileceğini göstermiştir [18]. Makalede, Support Vector Machine, Logistic Regression, Multilayer Perceptron Classifier, Decision Tree ve Random Forest algoritmaları kullanılmıştır [19]. Bu algoritmaların değerlendirme sonuçları Tablo 2’de verilmiştir.

Tablo 2. Choudhury A’nın makalesinin değerlendirme sonuçları

Algorithm	Accuracy	Recall	Precision
Logistic Regression	98.49	99.56	97.00
Decision Tree	97.95	96.98	98.22
Support Vector Classifier	97.30	97.99	95.82
Random Forest Classifier	98.14	98.49	97.20
MultilayerPerceptron Clas.	99.41	98.93	99.68

Bu makalede müşteri ve ürün bazında satın alma davranışlarını daha iyi anlamak için XAI kullanılacaktır. Kampanyaların ve diğer dış faktörlerin satın alma davranışlarını nasıl etkilediği veya bir ürünün satış miktarının hangi durumlarda değiştiğini gözlemlenecektir. Ayrıca AYZ tekniklerinin kullanımıyla elde edilen sonuçları, anlaşılabilir cümleler halinde sunarak kullanan kişilere rehber olması sağlanacak. Bulunan satın alma örüntüleri sayesinde de bir müşterinin hangi ürünlere ne sıklıkla ilgisi olduğu öğrenilebilir ve bu sayede işletmeler kampanyalarını optimize edebilir.

2 VERİ TOPLAMA VE SATIN ALMA DAVRANIŞLARI

Müşteri satın alma davranışları, işletmelerin ürün ve hizmetlerini daha iyi şekillendirebilmesi ve müşteri memnuniyetini arttırabilmesi için anlamaları gereken temel bir unsurdur. Müşterilerin neden belirli ürünleri tercih ettikleri, alışveriş sıklığı, harcama alışkanlıkları ve diğer davranışları, işletmelere rekabet avantajı sağlamak için kullanılabilir veri kaynaklarıdır. Bu davranışları anlamak, pazarlama stratejilerini optimize etmek ve müşteriye özel deneyimler sunmak için kritik bir adımdır [20]. Ayrıca bazı satın alma örüntülerini bulmak işletmelere paha biçilemez bilgiler sunabilir.

Müşteri satın alma örüntüleri, müşterilerin belirli bir süre içinde satın aldıkları ürünler, hizmetler veya deneyimler hakkındaki kalıplardır. Bu örüntüler, işletmelerin etkili satış stratejileri geliştirmesini, satışlarını arttırabilmesini ve müşterilerini daha iyi anlamasını sağlayabilir.

Müşteri satın alma örüntüleri, çeşitli faktörlerden etkilenebilir. Bu faktörler arasında demografik özellikler, satın alma geçmişi, ilgi alanları, davranışsal faktörler ve promosyonlar yer alabilir. Bu örüntüleri analiz etmek için çeşitli yöntemler kullanılabilir. Bu yöntemler arasında istatistiksel analiz, makine öğrenimi ve veri görselleştirme yer alır.

2.1 Veri İşlemleri

Müşteri satın alma davranışlarını incelemek için gerekli verilerin toplanması önemli bir adımdır. Bu çalışmada Kaggle adındaki veri bilimi platformunda bulunan “Retail Dataset” [21] adlı veri kümesi kullanılmıştır. Yaklaşık 3 yıllık veri bulunduran veri kümesinde, yaklaşık 29 bin veri ve 8 adet değişken bulunmaktadır. Tablo 3’te veri kümesinin ilk 5 verisi örnek verilmiştir.

Tablo 3. Veri kümesinin ilk 5 verisi

#	Unnamed:0	DocumentID	Date	SKU	Price	Discount	Customer	Quantity
0	0	716	2019-09-23	1039	381.78	67.37254	1	1.0
1	1	716	2019-09-23	853	593.22	0.00034	1	1.0
2	2	716	2019-09-23	862	423.73	-0.00119	1	1.0
3	3	716	2019-09-23	868	201.70	35.58814	1	1.0
4	4	716	2019-09-23	2313	345.76	61.01966	1	1.0

Tablo 3 incelendiğinde “Unnamed: 0” isimli kolona ihtiyaç olmadığı görülmüştür ve 100 defadan az satılan ürünler hataya yol açabileceğinden dolayı bu veriler de silinmiştir. Veri kümesinden daha çok çıkarım yapmak için veri kümesinin istatistiksel açıklaması Tablo 4’te yapılmıştır.

Tablo 4. Veri sütunlarının istatistik bilgileri

#	DocumentID	SKU	Price	Discount	CustomerID	Quantity
count	33356.00	33356.00	33356.00	33356.00	33356.00	33356.00
mean	8227.57	1241.27	2915.71	591.24	307.96	5.76
std	4028.04	893.87	10285.51	3508.98	162.67	19.55
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
max	15751.00	2904.00	522767.34	240396.66	605.00	1246.00

Toplamda 1939 farklı çeşitte 29103 ürün bulunuyor ve 14078 adet fatura kesilmiş. Bu durum, geniş bir ürün yelpazesi ve işlem hacmi olduğunu gösteriyor. Ortalama birim fiyat 2552 birim olmasına rağmen, minimum satış tutarının 0 olması dikkat çekiyor. Bu durumda temizlenmesi gereken veriler veya hatalı kayıtlar olabilir. İşlem sayısının müşteri sayısına oranının fazla olması, belirli ürünlere veya markalara sadık müşteri grupları olduğunu gösteriyor olabilir. Bu, müşteri sadakatinin yüksek olduğunu ve belirli ürünlerdeki talebin yüksek olduğunu gösterebilir. Ayrıca, iskonto miktarının oranı da bu sadakati etkileyebilir.

Veriyi daha anlamlı hale getirmek ve müşterilerin satın alma davranışlarını açıklayabilmek için indirim oranları anlamına gelen “Discount Percentage” (İskonto Oranı) isimli değişken eklenmiştir. Bu değişken eklenirken iskonto miktarı ürün fiyatına bölünerek 100 ile çarpılmıştır. Ayrıca “Effective Discount” (Etkili İskonto), “Applied Discount” (Uygulanan İskonto), “Total Discount (Toplam İndirim)”, “Average Discount Per Price” (Fiyat Başına Ortalama İndirim) adında dört değişken daha eklenmiştir. Quantity değişkeni yüksek, orta ve düşük isimli 3 değişkene bölünmüştür. Eğer satış miktarı

10'dan büyükse "Yüksek", 5'ten büyükse "Orta", 0 ile 5 arasında ise "Düşük" olarak segmente edilmiştir. Effective Discount değişkeni için indirim miktarı ürün fiyatından çıkarılmıştır ve gerçekleşen ödemenin miktarı bulunmuştur. Applied Discount değişkeni iskonto uygulanıp uygulanmadığını göstermektedir. Total Discount değişkeni satış başına uygulanan indirim miktarı ile satış adedi çarpımından elde edilebilir. Bu, toplam indirimi ifade eder. Average Discount Per Price değişkeni ise indirim miktarını fiyatla bölerek, her bir ürün için fiyat başına ortalama indirimi ifade eder. Ayrıca Date isimli değişken kullanılarak "Year" (yıl), "Month"(ay), "Day"(gün) ve "isWeekend" (hafta sonu mu?) isimli değişkenler türetilmiştir. Total Discount (Toplam İndirim): Satış başına uygulanan indirim miktarı ile satış adedi çarpımından elde edilebilir. Bu, toplam indirimi ifade eder. Uygulanan bu değişikliklerle veri kümesinin yeni hali Tablo 5'te gösterilmiştir.

Tablo 5. Düzenlenmiş veri kümesinin değişkenleri

Değişken Adı	Açıklama
DocumentID	İşlemin kimliği
SKU	Ürün kodu
Price	İşlemin satış fiyatı
Discount	İşlemin indirim tutarı (iskonto)
Customer	Müşteri kimlik numarası
Quantity	İşleminde satılan ürün sayısı
Discount Percentage	İndirim oranı
Effective Discount	Etkili İndirim
Applied Discount	İndirim uygulandı mı?
Year	İşlemin yılı
Month	İşlemin ayı
Day	İşlemin günü
isWeekend	Hafta sonu mu?
Total Discount	Toplam indirim
Average Discount Per Price	Fiyat başına ortalama indirim

3 MAKİNE ÖĞRENMESİ VE AÇIKLANABİLİR YAPAY ZEKA

Makine öğrenmesi, bilgisayarların verilere dayalı tahminler yapmak ve bunları iyileştirmek için kullandığı bir dizi yöntemdir [22]. Bu yöntemler, algoritmaları ve istatistiksel modelleri kullanarak veri setlerini inceleyip desenleri ve ilişkileri belirlemeyi sağlar. Makine öğrenimi süreci, performansı değerlendirmek için eğitim ve test olarak ikiye bölünür. Makine öğrenmesinin temel kavramları arasında sınıflandırma, regresyon ve kümeleme gibi yöntemler yer alır.

3.1 Açıklanabilir Yapay Zekâ Nedir?

Açıklanabilir Yapay Zekâ (AYZ), makine öğrenmesi modellerinin kararlarını anlaşılabilir bir şekilde sunmayı amaçlar. Makine öğrenmesi algoritmaları iç işleyişinin tamamen anlaşılmasını sebebiyle genellikle "siyah kutu" olarak kabul edilir. AYZ sayesinde model kararları yorumlanır ve bu sayede modelin güvenilirliği artırılabilir [3].

3.2 AYZ'nin Müşteri Davranışlarındaki Rolü

Müşteri davranışlarını incelemek için makine öğrenmesi ve AYZ'nin birleştirilmesi, işletmelere detaylı analizler sunabilir. Makine öğrenmesi modelleri, genellikle kararlarını belirli özelliklere dayandırır ve bu kararların neden ve nasıl alındığı konusunu açıklamada yetersiz olabilir. AYZ teknikleri, modelin o tahmini nasıl aldığını ve hangi özelliklerin tahminde belirleyici olduğunu açıklar [4].

Özellikle müşteri satın alma davranışlarını analiz ederken, AYZ'nin rolü kritik olabilir. Bu analizler, hangi faktörlerin müşteri kararlarını etkilediğini belirlemek için kullanılabilir. Örneğin, bir müşterinin belirli bir ürünü satın alıp almama kararı, ürünün fiyatı ve indirimler gibi faktörlerden etkilenebilir. AYZ, bu faktörlerin hangi ağırlıklara sahip olduğunu ve sonuçları nasıl etkilediğini anlamamızı sağlar.

3.3 Açıklanabilirlik Teknikleri

Açıklanabilirlik teknikleri, makine öğrenmesi modellerinin iç işleyişini ve sonuçlarını anlaşılır bir şekilde açıklamak amacıyla geliştirilmiş tekniklerdir.

3.3.1 Karar Ağaçları ve Açıklanabilirlik

Karar ağaçları, temel bir açıklanabilirlik tekniğidir. Veri kümesini sırayla bölerek ve belirli özelliklerin değerlerine dayanarak sonuçlara ulaşan bir ağaç yapısı kullanılır ve öğrenme işlemi bu şekilde gerçekleştirirler. Bu ağaçlar, karar sürecini adım adım ve özelliklerin önem sırasına göre analiz ederek sonuçları açıklar. Müşteri satın alma davranışlarını incelemek için karar ağaçları, müşterilerin davranışlarını daha anlaşılır bir şekilde açıklamak için etkili bir araç olarak kullanılabilir.

3.3.2 LIME (Local Interpretable Model-agnostic Explanations)

LIME, modelden bağımsız lokal açıklamaların oluşturulmasını sağlayan bir açıklanabilirlik yöntemidir. Bu metodoloji, belirli bir örneği seçer ve bu örneğin tahminini yerel bir model ile yaklaşık olarak açıklar. Müşteri satın alma davranışlarını analiz etmek için LIME, belirli bir müşterinin belirli bir kararını neden verdiğini anlamak için örnek bazlı açıklamalar yaparak analizin daha iyi anlaşılmasına olanak tanır.

3.3.3 SHAP (SHapley Additive exPlanations)

SHAP, kooperatif oyun teorisinden esinlenerek geliştirilmiş bir açıklanabilirlik tekniğidir. Bu teknik, her bir özelliğin model tahminine nasıl katkıda bulunduğunu ölçer ve gösterir. Müşteri satın alma davranışlarını analiz etmek için SHAP, hangi özelliklerin müşteri kararlarını ne ölçüde etkilediğini daha ayrıntılı bir şekilde anlamamızı sağlar.

4 MAKİNE ÖĞRENMESİ MODELİNİN OLUŞTURULMASI

Bu projede kullanmak için makine öğrenmesi algoritmalarından Random Forest Classifier, XGBoost ve CatBoost seçilmiştir. Kullanılan makine öğrenmesi algoritmalarını açıklamak için ise SHAP kullanılacaktır.

- Random Forest Classifier: Çok sayıda karar ağacının bir araya gelerek oluşturduğu bir topluluk öğrenme algoritmasıdır.
- XGBoost (Extreme Gradient Boosting): Gradient Boosting algoritmasının hızlandırılmış ve iyileştirilmiş bir sürümüdür. Gradient Boosting'den farklı olarak, hata hesaplamalarını optimize eder ve ağaçları sırayla oluşturur.
- CatBoost: Özellikle kategorik değişkenleri işleme yeteneği ile bilinen bir Gradient Boosting algoritmasıdır. Kategorik değişkenleri otomatik olarak işler, böylece veriyi dönüştürme veya kodlama gerekliliği azalır.

4.1 Değerlendirme Metrikleri

Seçilen algoritmalarla eğitim verileri üzerinde model oluşturma süreci, genellikle iteratif bir yaklaşımla gerçekleştirilir. Eğitim verileri üzerinde model eğitilir ve ardından test verileri üzerinde değerlendirilir. Değerlendirme metrikleri, modelin performansını anlamak için kullanılır. Örnek metrikler şunlar olabilir:

- Doğruluk (Accuracy): Doğru tahminlerin toplam örnek sayısına oranıdır.
- Hassasiyet (Precision): Pozitif olarak tahmin edilenlerden gerçekten pozitif olanların oranını belirtir.
- Duyarlılık (Recall): Gerçek pozitiflerin tespit edilen pozitiflerin oranını ifade eder.
- F1-Skoru: Hassasiyet ve duyarlılık arasındaki dengeyi gösteren bir metriktir.

5 MÜŞTERİ DAVRANIŞLARININ ANALİZİ VE YORUMLANMASI

Müşteri satın alma davranışlarını analiz etmek, satın alma örüntülerini bulmak ve hangi durumların (indirim, fiyat) müşteri satışları üzerinde etkili olduğunu anlamak, işletmeler için paha biçilmez bir görüş sağlar. Bu aşamada, oluşturulan modeli kullanarak verileri incelemek ve müşteri davranışlarına dair değerli görüşler elde etmek kritik öneme sahiptir. Bu bölümde, model sonuçlarını yorumlamak ve analiz etmek için izlenebilecek adımları ele alacağız.

5.1 Müşteri Bazında Satın Alma Örüntüleri

Veriler incelendiğinde, müşterilerin satın alma davranışlarında bazı örüntüler olduğu ortaya çıkmaktadır. Bu örüntüler, müşterilerin ihtiyaçlarını, ilgi alanlarını ve satın alma kararlarını anlamamıza yardımcı olabilir. Veri kümesinden çıkarılan örnek müşteri davranışları aşağıdaki tablolarda verilmiştir.

Tablo 6. Müşteri 75'in satın alma davranışları

Müşteri numarası (ID)	Alışkanlıkları
75	En çok satılan ürünlerden alıyor, düşük alım sıklığına sahip, yüksek fiyatları alma eğiliminde, ayın ortasında satın alıyor, yüksek harcama yapma eğiliminde, indirim olunca alışveriş yapıyor. En çok satın aldığı SKU: 1443, 1408, 978 ve en çok alışverişi yılın eylül ayında yapmaktadır.

Tablo 7. Müşteri 167'un satın alma davranışları

Müşteri numarası (ID)	Alışkanlıkları
167	En az satılan ürünleri alıyor, yüksek alım sıklığına sahip, yüksek fiyatları alma eğiliminde, ayın başında satın alıyor, düşük harcama yapma eğiliminde, indirim olunca alışveriş yapıyor. En çok satın aldığı SKU: 51, 543 ve en çok alışverişi yılın nisan ayında yapmaktadır.

Tablo 8. Müşteri 434'ün satın alma davranışları

Müşteri numarası (ID)	Alışkanlıkları
434	En çok satılan ürünlerden alıyor, orta alım sıklığına sahip, yüksek fiyatları alma eğiliminde, ayın ortasında satın alıyor, yüksek harcama yapma eğiliminde, indirimleri önemsemeden alışveriş yapıyor. En çok satın aldığı SKU: 1345, 1346, 715 ve en çok alışverişi yılın ekim ayında yapmaktadır.

Tablo 9. Müşteri 455'in satın alma davranışları

Müşteri numarası (ID)	Alışkanlıkları
455	En çok satılan ürünlerden alıyor, yüksek alım sıklığına sahip, yüksek fiyatları alma eğiliminde, ayın başında satın alıyor, yüksek harcama yapma eğiliminde, indirimleri önemsemeden

alışveriş yapıyor. En çok satın aldığı SKU: 2649, 2817, 501 ve en çok alışverişini yılın mart ayında yapmaktadır.

Tablo 10. Müşteri 495'in satın alma davranışları

Müşteri numarası (ID)	Alışkanlıkları
495	En az satılan ürünleri alıyor, düşük alım sıklığına sahip, düşük fiyatları alma eğiliminde, ayın ortasında satın alıyor, yüksek harcama yapma eğiliminde, indirim olunca alışveriş yapıyor. En çok satın aldığı SKU: 2900 ve en çok alışverişini yılın ağustos ayında yapmaktadır.

Tablo 11. Müşteri 598'in satın alma davranışları

Müşteri numarası (ID)	Alışkanlıkları
598	En çok satılan ürünlerden alıyor, düşük alım sıklığına sahip, yüksek fiyatları alma eğiliminde, ayın sonunda satın alıyor, yüksek harcama yapma eğiliminde, indirim olunca alışveriş yapıyor. En çok satın aldığı SKU: 166, 272, 541 ve en çok alışverişini yılın ocak ayında yapmaktadır.

Müşterilerin satın alma geçmişlerine bakarak her müşterinin potansiyel farklı bir alışkanlık çıktısı olabilir. Bunlar toplam 216 farklı şekilde oluşabilir, bu çıktılara ek olarak müşterilerin en çok satın aldığı SKU'lar ve en çok alışveriş yaptıkları ayın bilgisi de eklenmiştir.

Müşterinin ürün tercihlerine göre “en çok satılan ürünlerden alıyor” ve “en az satılan ürünleri alıyor” şeklinde 2 farklı potansiyel durum oluşmaktadır.

Müşterinin alım sıklığına göre “düşük alım sıklığına sahip”, “orta alım sıklığına sahip” ve “yüksek alım sıklığına sahip” şeklinde 3 farklı potansiyel durum oluşmaktadır.

Müşterinin fiyat tercihlerine göre “düşük fiyatları alma eğiliminde”, “orta fiyatları alma eğiliminde” ve “yüksek fiyatları alma eğiliminde” şeklinde 3 farklı potansiyel durum oluşmaktadır.

Müşterinin ayın hangi bölümünde daha çok alışveriş yaptığına göre “ayın başında satın alıyor”, “ayın ortasında satın alıyor” ve “ayın sonunda satın alıyor” şeklinde 3 farklı durum oluşmaktadır.

Müşterinin yaptığı toplam harcamalara göre “yüksek harcama yapma eğiliminde” ve “düşük harcama yapma eğiliminde” şeklinde 2 farklı potansiyel durum oluşmaktadır.

Müşterinin indirim durumlarına tepkisine göre “indirim olunca alışveriş yapıyor” ve “indirimleri önemsemeyen alışveriş yapıyor” şeklinde 2 farklı potansiyel durum oluşmaktadır.

Ek olarak, müşterinin en çok satın aldığı SKU'lar ve en çok alışveriş yaptığı ay çıktıya eklenmiştir. Toplamda 216 farklı kombinasyon oluşturulmuştur. Müşterinin alışkanlıklarına göre yukarıda yazılanları içeren bir çıktı oluşturulmaktadır. Farklı çıktı örnekleri aşağıda verilmiştir:

- 1 “En az satılan ürünleri alıyor, yüksek alım sıklığına sahip, yüksek fiyatları alma eğiliminde, ayın ortasında satın alıyor, yüksek harcama yapma eğiliminde, indirim olunca alışveriş yapıyor”
- 2 “En çok satılan ürünlerden alıyor, yüksek alım sıklığına sahip, yüksek fiyatları alma eğiliminde, ayın ortasında satın alıyor, yüksek harcama yapma eğiliminde, indirim olunca alışveriş yapıyor”
- 3 “En az satılan ürünleri alıyor, düşük alım sıklığına sahip, düşük fiyatları alma eğiliminde, ayın başında satın alıyor, yüksek harcama yapma eğiliminde, indirim olunca alışveriş yapıyor”

- 4 “En çok satılan ürünleri alıyor, orta alım sıklığına sahip, yüksek fiyatları alma eğiliminde, ayın başında satın alıyor, yüksek harcama yapma eğiliminde, indirim olunca alışveriş yapıyor”
- 5 “En az satılan ürünleri alıyor, yüksek alım sıklığına sahip, yüksek fiyatları alma eğiliminde, ayın ortasında satın alıyor, yüksek harcama yapma eğiliminde, indirim olunca alışveriş yapıyor”
- 6 “En az satılan ürünleri alıyor, düşük alım sıklığına sahip, düşük fiyatları alma eğiliminde, ayın başında satın alıyor, düşük harcama yapma eğiliminde, indirim olunca alışveriş yapıyor”
- 7 “En çok satılan ürünleri alıyor, yüksek alım sıklığına sahip, orta fiyatları alma eğiliminde, ayın başında satın alıyor, düşük harcama yapma eğiliminde, indirim olunca alışveriş yapıyor”
- 8 “En çok satılan ürünleri alıyor, yüksek alım sıklığına sahip, yüksek fiyatları alma eğiliminde, ayın ortasında satın alıyor, yüksek harcama yapma eğiliminde, indirim olunca alışveriş yapıyor”
- 9 “En az satılan ürünleri alıyor, düşük alım sıklığına sahip, düşük fiyatları alma eğiliminde, ayın başında satın alıyor, yüksek harcama yapma eğiliminde, indirim olunca alışveriş yapıyor”
- 10 “En çok satılan ürünleri alıyor, yüksek alım sıklığına sahip, yüksek fiyatları alma eğiliminde, ayın ortasında satın alıyor, düşük harcama yapma eğiliminde, indirim olunca alışveriş yapıyor”
- 11 “En az satılan ürünleri alıyor, düşük alım sıklığına sahip, yüksek fiyatları alma eğiliminde, ayın sonunda satın alıyor, yüksek harcama yapma eğiliminde, indirim olunca alışveriş yapıyor”
- 12 “En az satılan ürünleri alıyor, düşük alım sıklığına sahip, düşük fiyatları alma eğiliminde, ayın başında satın alıyor, düşük harcama yapma eğiliminde, indirim olunca alışveriş yapıyor”
- 13 “En çok satılan ürünleri alıyor, düşük alım sıklığına sahip, düşük fiyatları alma eğiliminde, ayın sonunda satın alıyor, yüksek harcama yapma eğiliminde, indirimleri önemsemeden alışveriş yapıyor”
- 14 “En çok satılan ürünlerden alıyor, yüksek alım sıklığına sahip, yüksek fiyatları alma eğiliminde, ayın ortasında satın alıyor, yüksek harcama yapma eğiliminde, indirim olunca alışveriş yapıyor”
- 15 “En az satılan ürünleri alıyor, düşük alım sıklığına sahip, düşük fiyatları alma eğiliminde, ayın sonunda satın alıyor, yüksek harcama yapma eğiliminde, indirimleri önemsemeden alışveriş yapıyor”

Müşterilerin ürün tercihlerine göre 2, alım sıklıklarına göre 3, fiyat tercihlerine göre 3, ayın hangi bölümünde satın aldıklarına göre 3, yaptıkları toplam harcamalara göre 2 ve indirim durumlarına göre 2 farklı örnek, toplamda 15 farklı örnek verilmiştir. Oluşabilecek tüm potansiyel farklı durumlara ait bölümlerin altı çizilmiştir.

6 MODEL DEĞERLENDİRMELERİ

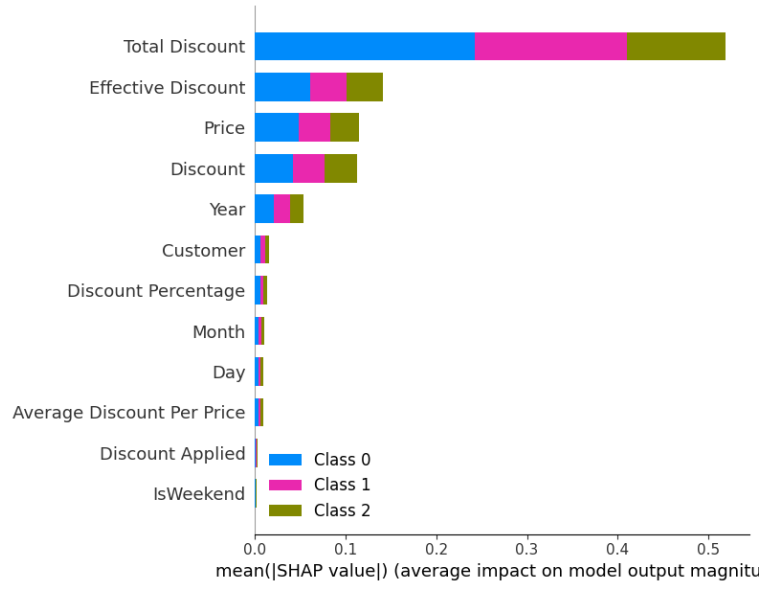
Makalede kullanılan modellerin değerlendirme metrikleri Tablo 12’de verilmiştir.

Tablo 12. Müşteri satın alma davranışları değerlendirme metrikleri

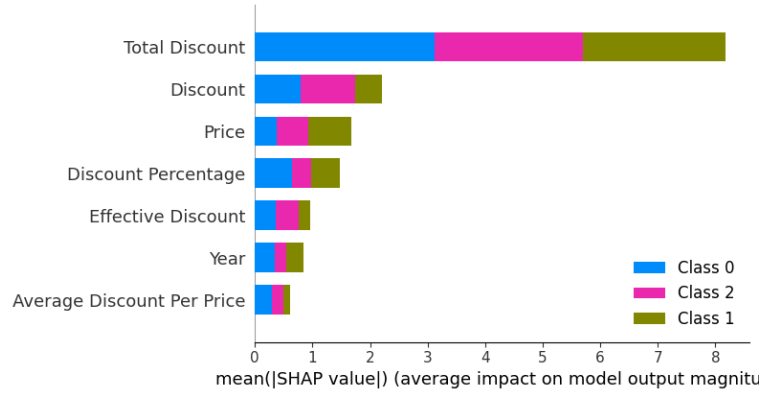
Model	Accuracy	Precision	Recall	F1-Score
Random Forest	0.9921	0.99	0.99	0.99
XGBoost	0.9904	0.98	0.98	0.98
CatBoost	0.9908	0.98	0.99	0.99

6.1 Açıklanabilirlik Tekniklerinin Uygulanması

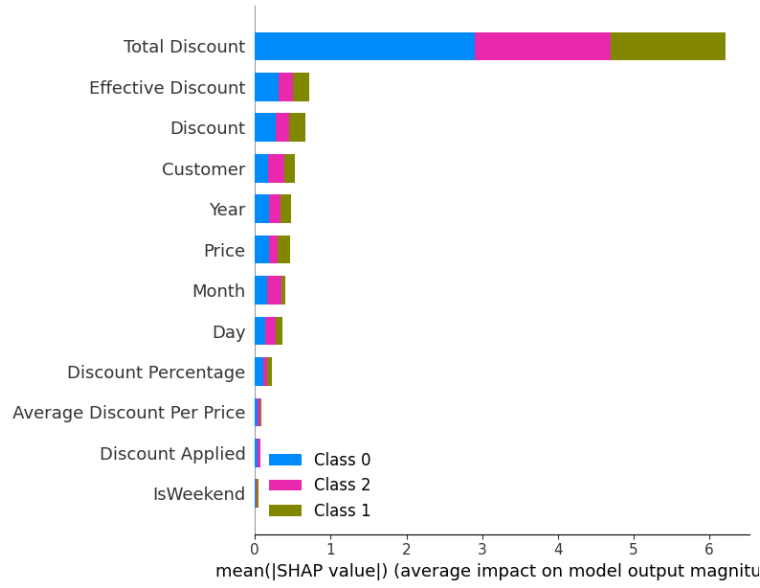
Açıklanabilirlik teknikleri, model sonuçlarını daha anlaşılır ve yorumlanabilir hale getirmek için kullanılır. Bu projede açıklanabilirlik tekniği olarak SHAP kullanılmıştır. SHAP görselleri aşağıdaki gibi Figür 1, Figür 2 ve Figür 3’te verilmiştir.



Figür 1: RandomForestClassifier algoritmasının SHAP değerleri



Figür 2: XGBoost algoritmasının SHAP değerleri



Figür 3: CatBoost algoritmasının SHAP değerleri

6.2 Model Sonuçlarının İncelenmesi ve Açıklanması

Bu bölümde, üç farklı makine öğrenmesi modeli olan Random Forest, XGBoost ve CatBoost kullanılarak elde edilen sonuçlar incelenecektir. Bu sonuçlar, müşteri satın alma davranışlarını anlamayı ve işletme stratejilerini optimize etmeyi sağlayabilir.

6.2.1 Random Forest Classifier Sonuçları

Random Forest modeli %99,21 doğruluk oranına sahiptir. Model her sınıf için yüksek precision, recall ve f1-score metriklerine sahiptir. Bu metrikler, modelin her sınıfı başarıyla sınıflandırdığını ve yanlış sınıflandırma oranlarının düşük olduğunu gösterir. Ayrıca, SHAP ile hesaplanan özellik önem sıralamalarına göre, müşteri satın alma davranışını etkileyen en önemli faktörler "Total Discount" ve "Effective Discount" olarak belirlenmiştir.

6.2.2 XGBoost Sonuçları

XGBoost modeli de yüksek bir doğruluk oranına sahiptir ve bu oran 99,04'tür. Model her sınıf için yüksek precision, recall ve f1-score metriklerine sahiptir. Bu sonuçlar, XGBoost'un güçlü bir performans sergilediğini göstermektedir. Ancak RandomForestClassifier'a göre geride kalmıştır. Ayrıca, SHAP ile hesaplanan özellik önem sıralamalarına göre, "Total Discount", "Discount" ve "Year" özellikleri müşteri satın alma davranışını etkileyen en önemli faktörler olarak öne çıkmaktadır.

6.2.3 CatBoost Sonuçları

CatBoost modeli, RandomForestClassifier'dan sonra en yüksek doğruluk oranının olan %99,08'e sahiptir. Değerlendirme metrikleri olan precision, recall ve f1-score değerleri de oldukça yüksektir. Bu da CatBoost'u RandomForestClassifier'dan sonra en başarılı model yapar. SHAP ile yapılan CatBoost özellik önem sıralamalarına göre, "Total Discount" değişkeni müşteri satın alma davranışını etkileyen en önemli faktör olarak belirlenmiştir.

7 TARTIŞMA

Bu çalışmada, üç farklı sınıflandırma modeli (Random Forest, XGBoost ve CatBoost) kullanılarak müşteri satın alma davranışları analiz edildi. Özet sonuçlar aşağıdaki gibidir:

- Random Forest Classifier: Precision, Recall ve F1-score değerleri, her üç sınıf için de yüksektir, bu da modelin genel performansının iyi olduğunu gösterir. SHAP analizi, "Total Discount" özelliğinin diğerlerine göre daha fazla öneme sahip olduğunu ortaya koyuyor.
- XGBoost: Bu model de yüksek bir doğruluk oranına sahiptir ve diğer metrikler açısından da tatmin edici sonuçlar vermektedir. "Discount" ve "Total Discount" özellikleri, modelin tahminlerinde önemli bir rol oynamaktadır.
- CatBoost: Bu model de başarılı bir şekilde sınıflandırma yapmaktadır. XGBoost'tan daha iyi değerlere sahiptir. SHAP analizi, "Total Discount" özelliğinin bu model için en belirgin belirleyici olduğunu gösteriyor

Açıklanabilirlik teknikleri bu projede kullanılan haliyle SHAP, her üç model için de özelliklerin önem sıralamalarını belirlemede etkili olmuştur. Bu teknikler, hangi özelliklerin müşteri satın alma davranışlarını en çok etkilediği konusunda bilgiler ve şekiller sunmuştur. Özellikle "Total Discount" özelliği, her üç model için de en önemli belirleyici olarak ortaya çıkmıştır. Bu, "Total Discount" faktörünün müşteri satın alma davranışları üzerindeki etkisinin yüksek olduğunu göstermektedir. Elde edilen modeller, işletmelere müşteriye özel pazarlama stratejileri geliştirmede yardımcı olabilir. Özellikle "Total Discount" gibi önemli özelliklerin kullanılması, pazarlama kampanyalarının etkisini artırabilir. Kullanılan modeller, müşteri taleplerini tahmin etmek için kullanılabilir ve bu sayede, ürün stoklarının daha verimli bir şekilde yönetilmesini sağlayabilir. Ayrıca bütçeyi optimize etmek ve etkili bir fiyatlandırma stratejisi oluşturmak için kullanılabilir. Bu kullanım alanları sayesinde, işletmelere müşteriye daha iyi anlama olanağı sağlanır ve bu sayede işletmeler rekabet avantajı elde edebilirler. Satın alma örüntülerinin bulunmasıyla müşterinin hangi ürünü ne zaman alacağı tahmin edilebilir, bu müşteriye özel öneri kampanyaları düzenlenebilir.

7 SONUÇ

Bu proje, müşteri satın alma davranışlarını anlamak ve tahmin etmek için satın alma örüntüleri bulunmuş, makine öğrenimi tekniklerini kullanmış ve açıklanabilirlik tekniklerini uygulamıştır. Bu süreçte, üç farklı sınıflandırma modeli kullanılmış ve çeşitli özelliklerin müşteri satın alma davranışları üzerindeki etkisi analiz edilmiştir. Satın alma örüntüleri bulunmuş, bu örüntüler için potansiyel 216 farklı çıktı oluşturulmuştur. Müşterinin satın alma davranışlarına göre nu potansiyel çıktılardan oluşturulacaktır. Ayrıca, model sonuçları açıklanabilir hale getirilmiş ve işletmelere değerli bilgiler sunulmuştur. Bu çalışma, müşteri satın alma davranışlarını etkileyen önemli faktörleri belirlemede ve bu davranışları tahmin etmede etkili bir yol sunmaktadır. Ana bulgular şunlardır: "Total Discount" gibi özellikler, müşteri satın alma davranışlarını etkileyen en önemli faktörler olarak belirlenmiştir. Modeller yüksek doğruluk oranlarına sahiptir ve her üç sınıf için de iyi bir performans sergilemektedirler. Açıklanabilirlik teknikleri, model sonuçlarını anlamada ve işletme stratejilerini optimize etmede önemli bir rol oynamıştır. Bu bulgular, işletmelerin müşteri davranışlarını daha iyi anlayarak kararlarını iyileştirmelerine yardımcı olabilir.

REFERANSLAR

- [1] Hibrit Açıklanabilir Yapay Zekâ Tasarımı ve LIME Uygulaması Ayça ÇAKMAK PEHLİVANLI, Rahmi Ahmet Selim DELİLOĞLU
- [2] Nagaraj, P., Muneeswaran, V., Dharamidharan, A., Balanathanan, K., Arunkumar, M., & Rajkumar, C. (2022, April). A Prediction and Recommendation System for Diabetes Mellitus using XAI-based Lime Explainer. In 2022 International Conference on Sustainable Computing and Data Communication Systems (ICSCDS) (pp. 1472-1478). IEEE.
- [3] Terzi, R. (2021). Yapay Zekâ ve Büyük Veri Çalışmaları, Siber Güvenlik ve Mahremiyet (Chapter 6)
- [4] Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence*, 1(5), 206-215

- [5] A historical perspective of explainable Artificial Intelligence, Roberto Confalonieri, Ludovik Coba, Benedikt Wagner, Tarek R. Besold Brunato, M., Battiti, R.
- [6] Kaya, B., & Alhajj, R. (2019). Açıklanabilir yapay zeka: Yeni bir yaklaşım ve bir uygulama. *ACM Bilişim Dergisi*, 30(1), 22-29
- [7] Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Giannotti, F., & Pedreschi, D. (2018). A survey of methods for explaining black box models. *ACM computing surveys (CSUR)*,
- [8] Çyras, K., Rago, A., Albini, E., Baroni, P., & Toni, F. (2021). Argumentative XAI: a survey.
- [9] İnan, A., Kocadag, M., & Şen, S. (2020). Türkiye'de yapay zeka uygulamaları: mevcut durum, fırsatlar ve tehditler. *Teknolojik Araştırmalar: Yeni Teknoloji, Yeni Toplum*, 1(1), 45-54
- [10] Chromik, M. (2020). reshape: A framework for interactive explanations in xai based on shap.
- [11] Mahbooba, B., Timilsina, M., Sahal, R., & Serrano, M. (2021). Explainable artificial intelligence (XAI) to enhance trust management in intrusion detection systems using decision tree model. *Complexity*, 2021, 1-11
- [12] Weber, L., Lapuschkin, S., Binder, A., & Samek, W. (2022). Beyond explaining: Opportunities and challenges of XAI-based model improvement. *Information Fusion*
- [13] Tuğçe Akın, "Müşteri Memnuniyetinin İncelenmesi", YBS Ansiklopedi, v.5, is. 2, pp.18- 25, 2018
- [14] Sadi Evren Seker, "Müşteri Kayıp Analizi", YBS Ansiklopedi, v.3, is. 1, pp. 26 – 28, 2016
- [15] Explainable AI for Credit Assessment in Banks by Petter Eilif de Lange, Borger Melsom, Christian Bakke Vennerød and Sjur Westgaard 2
- [16] Augmented Cross-Selling Through Explainable AI- A Case from Energy Retailing Molnar, C. *Interpretable Machine Learning*. 2019
- [17] Machine Learning Based Prediction of Consumer Purchasing Decisions: The Evidence and Its Significance Saavi Stubseid and Ognjen Arandjelovic
- [18] A Machine Learning Approach to Identify Potential Customer Based on Purchase Behavior Adil Mahmud Choudhury
- [19] Seker, Sadi Evren, and Ibrahim Ocak. "Performance prediction of roadheaders using ensemble machine learning techniques." *Neural Computing and Applications* 31.4 (2019): 1103-1116.
- [20] Familmaleki M., Aghighi A., Analyzing the Influence of Sales Promotion on Customer Purchasing Behavior
- [21] Şadi Evren ŞEKER, "Retail Data Set", Mayıs 2023. from <https://www.kaggle.com/datasets/shedai/retail-data-set>
- [22] Molnar, C. (2021). Interpretable machine learning: A guide for making black box models explainable. Chapter: What Is Machine Learning?