



## DEMOGRAFİK ÖZELLİKLERİN ÇEVİRİMİÇİ MARKET ALIŞVERİŞİ KULLANIMINA ETKİSİNİN MAKİNE ÖĞRENMESİ YÖNTEMLERİ İLE TAHMİNİ

<sup>1</sup> Burak BAHÇIVAN , <sup>2,\*</sup> Atınc YILMAZ 

<sup>1</sup>İstanbul Beykent Üniversitesi, Lisansüstü Eğitim Enstitüsü, Bilgisayar Mühendisliği Bölümü, İstanbul,  
TÜRKİYE

<sup>2</sup>İstanbul Beykent Üniversitesi, Mühendislik-Mimarlık Fakültesi, Bilgisayar Mühendisliği Bölümü, İstanbul,  
TÜRKİYE

<sup>1</sup>[burakbahcivan@gmail.com](mailto:burakbahcivan@gmail.com), <sup>2</sup>[atincyilmaz@beykent.edu.tr](mailto:atincyilmaz@beykent.edu.tr)

### Önemli Katkılar (Highlights)

- Genellikle performans ölçümü ve bankacılık sektöründe tercih edilen skorlama Lojistik Regresyon algoritması gauss dağılımı ile birlikte kullanılmıştır.
- Sık kullanılan makine öğrenmesi yöntemleri kullanılarak bir kullanıcının bir kategoriden alışveriş yapma olasılığını öngörülmesi ve bu tercihi etkileyen demografik özellikler üzerine odaklanılmıştır
- Çalışmada Lojistik Regresyon, Karar Ağacı, K-En Yakın Komşu, Gradyan Arttırılmış Ağaçlar ve Rastgele Orman yöntemleri uygulanmıştır.
- Çevrimiçi market uygulaması ya da web hizmeti sunan tüm firmaların fayda sağlayabilmesi adına herhangi bir kategoriden alışveriş yapabilecek kullanıcının sahip olması gereken özellikler ortaya çıkartılabilmekte; bu sayede kişiselleştirilmiş tahmin sistemi tasarlanabilmektedir.



## DEMOGRAFİK ÖZELLİKLERİN ÇEVİRİMİÇİ MARKET ALIŞVERİŞİ KULLANIMINA ETKİSİNİN MAKİNE ÖĞRENMESİ YÖNTEMLERİ İLE TAHMİNİ

<sup>1</sup>Burak BAHÇIVAN , <sup>2,\*</sup>Atınc YILMAZ 

<sup>1</sup>*İstanbul Beykent Üniversitesi, Lisansüstü Eğitim Enstitüsü, Bilgisayar Mühendisliği Bölümü, İstanbul, TÜRKİYE*

<sup>2</sup>*İstanbul Beykent Üniversitesi, Mühendislik-Mimarlık Fakültesi, Bilgisayar Mühendisliği Bölümü, İstanbul, TÜRKİYE*

<sup>1</sup>[burakbahcivan@gmail.com](mailto:burakbahcivan@gmail.com), <sup>2</sup>[atincyilmaz@beykent.edu.tr](mailto:atincyilmaz@beykent.edu.tr)

(Geliş/Received: 10.07.2022; Kabul/Accepted in Revised Form: 07.08.2023)

**ÖZ:** Çevrimiçi market alışverişi hizmeti sağlayan firmaların ürün satışlarını arttırmak ve yeni müşteriler elde etmek amacıyla hayata geçirdikleri birçok kampanyanın başarıya ulaşamadığı görülmektedir. Çevrimiçi alışverişte ürün satışlarının artmasını ve kampanyaların başarılı olmasını amaçlayarak, çevrimiçi market üzerinden alışveriş yapan 394 kullanıcıya ait çeşitli veriler anket aracılığı ile toplanmıştır. Çalışmada, sık kullanılan makine öğrenmesi algoritmaları ile modelleme yapılarak çevrimiçi market kullanıcılarının kişisel bakım kategorisinden alışveriş yapma eğilimlerinin öngörülmesini sağlayan bir model ortaya konulmuştur. Karar Ağaçları, K-En Yakın Komşu, Gradyan Arttırılmış Ağaçlar, Rastgele Orman ve Lojistik Regresyon yöntemleri modelleme için kullanılmıştır. Son olarak da ortaya çıkan eğri altında kalan alan (EAKA-AUC), geri çağırma (recall), f<sub>1</sub>-skor (f<sub>1</sub>-score) değerleri üzerinden yöntemlerin performans karşılaştırılması yapılmıştır. Çalışma sonucunda en yüksek performans 0.928 doğruluk oranı ve 0.92 AUC değerleri ile rastgele orman yöntemi ile elde edilirken; Gradyan Arttırılmış Ağaçlar yöntemi uygulanan model ise 0.704 doğruluk oranı ve 0.70 AUC değeri ile en düşük performansa ulaşmıştır. Çalışmada elde edilen bulgulara göre, özellikleri "43-47 yaş altı, günlük internet kullanımı fazla, kapıda kredi kartı ile ödeme tercihi yapmayan" kullanıcıların kişisel bakım kategorisinden alışveriş yapmayı tercih ettikleri analiz edilmiştir. Elde edilen sonuçlar incelendiğinde, şirketlerin çalışmada ortaya konan model ile hedef müşteri kitlelerini daha iyi belirlemelerine olanak sağlanacağı ve bu sayede yapılan gereksiz yatırımların önüne geçilebileceği öngörülmektedir.

**Anahtar Kelimeler:** Makine Öğrenmesi, Sınıflandırma, Skorlama, Çevrimiçi Alışveriş Tahmini

### Prediction of The Effect of Demographic Features on Online Market Shopping Using Machine Learning Methods

**ABSTRACT:** It is seen that many of the campaigns implemented by companies providing online shopping services to increase their product sales and acquire new customers have not been successful. In order to increase product sales in online shopping and to make campaigns successful, various data of 394 users who shopped on the online market were collected through a survey. In the study, a model that provides prediction of the shopping tendencies of online market users from the personal care category has been presented by applying frequently used machine learning algorithms. Decision Trees, K-Nearest Neighbor, Gradient Boosted Trees, Random Forest and Logistic Regression methods were used for modeling. In addition, analysis has been done with AUC (Area under the Curve), recall, f<sub>1</sub>-score values to compare the performance of the applied models. As a result of the study, the highest performance was obtained with the random forest method with an accuracy rate of 0.928 and AUC of 0.92. The model which applied with Gradient Boosted Trees method was achieved the lowest performance values with an accuracy rate of 0.704 and an AUC value of 0.70. According to the findings in the study, it has been analyzed that users whose properties are "under the age of 43-47, high daily internet user, do not choice to pay by credit card at the door" prefer shopping from the personal care category. When the results

obtained are examined, it is predicted that the model set forth in the study will enable companies to determine their target customer groups better, and thus, unnecessary investments can be prevented.

**Keywords:** *Machine Learning, Classification, Scoring, Online Shopping Prediction*

## 1. GİRİŞ (INTRODUCTION)

Teknolojinin hızla gelişmesi beraberinde tüketimin teknoloji ile olan etkileşimini arttırmaktadır. Günümüzde birçok firma çevrimiçi satış hizmetini müşterilerine sağlamak adına çeşitli yatırımlar yapmaktadır. Bu yatırımlar, ürün satışlarını arttırmak, yeni müşteri elde etmek, stok miktarını azaltma ve kârlılığı arttırma gibi amaçlar doğrultusunda yapılmaktadır. Ancak bu kampanyaların bir kısmının hedefine ulaşmadığı, kullanıcılar tarafından kullanılmadığı görülmektedir. Bu da şirketlerin kampanyalar için yaptığı yatırımların istenilen faydayı sağlamadığını ortaya çıkartmaktadır. Bu çalışma ile var olan bu sorun ele alınarak, müşterilerden elde edilecek olan demografik bilgiler ile kişisel bakım kategorisinden alışverişin tercih durumuna ilişkin bir öngörü oluşturmak amaçlanmaktadır. Bu sayede etkin bir kampanya stratejisi için şirketlerin faydalanabileceği bir model oluşturmak ve alanda devam eden çalışmalara katkı sağlanması düşünülmektedir.

Genellikle performans ölçümleri ve bankacılık sektöründeki kredi risk skorlaması gibi alanlarda tercih edilen skorlama için lojistik regresyon algoritması ile gauss dağılımı çalışmada birlikte kullanılmıştır. Uygulamada kullanılan veri seti, rastgele seçilen 394 çevrimiçi market kullanıcısının demografik ve çevrimiçi market alışveriş tecrübesine ait bilgilerin anket aracılığı ile elde edilmesi ile oluşturulmuştur. Toplamam bu veriler ile model oluşturulmuş ve kullanıcıların "Kişisel Bakım" kategorisinden alışveriş yapmayı tercih etme durumu öngörülmeye çalışılmıştır. Ayrıca, çalışma sırasında kullanıcıların hangi demografik özelliklerinin etken olduğu tespit edilmesi amaçlanmıştır. Çalışmanın şirketlerin yapacağı gereksiz yatırımların önüne geçmesi, hangi kullanıcılara bu kategoriden kampanya sunması gerektiğine dair bir fikir sağlaması hedeflenmiştir [1].

Wang ve ark., 2022 yılında çevrimiçi alışveriş davranışlarını baz alarak kronik hastalıklarının tahmini için öğrenme tabanlı modelleme kullanmışlardır [2]. Çalışmada kronik hastalık olarak diyabet ve depresyon öngörüsü hedeflenmiş ve önerdikleri öğrenme tabanlı model ile yüksek doğruluk elde etmişlerdir. Moon ve ark., 2021 yılında müşterilerin çevrimiçi platformlarda alışveriş davranışlarını makine öğrenmesi algoritmaları ile analiz ederek müşteri memnuniyetini arttırmayı amaçlamışlardır [3]. Çalışmada Naif Bayes, Apriori, Karar Ağacı ve Rastgele Orman yöntemleri uygulanmış; en yüksek doğruluğa Apriori algoritması modeli ile %88 ve Naif Bayes modeli ile %87 olarak elde edilmiştir. Koehn ve ark., 2020 yılında derin öğrenme yöntemi ile kullanıcıların tıklama verilerini kullanarak alışveriş davranışlarını tahmin etmişlerdir [4]. Çalışmada kapalı tekrarlayan birime sahip derin sinir ağları, uzun kısa süreli bellek modelleri ve tekrarlayan sinir ağları modelleri kullanılmıştır. Wang ve ark., 2022 yılında makine öğrenmesi algoritmalarını kullanarak çevrimiçi konut reklamlarının analizi ile kentsel sosyoekonomik durum analizi üzerinde çalışmışlardır [5]. Çalışmada yedi farklı makine öğrenmesi algoritması modelleme için kullanılmış; ağaç tabanlı algoritmaların uzamsal tahmin için daha güçlü sonuçlar elde ettiği gözlenmiştir. Volkmar ve ark., 2022 yılında yaptıkları çalışmada pazarlama yönetimi alanında yapay zeka algoritmaları ile geleceğe yönelik engellerin ve gelişmelerin öngörülmesini amaçlamışlardır [6]. Çalışma neticesinde yapılan anket çalışması sonucunda karar verme, müşteri yönetimi ve strateji oluşturma ve uygulama konusunda organizasyonların yaşadığı zorluklar olarak gruplandırıldığı belirtilmiştir. Salminen ve ark., 2019 yılında içerik yönetimi konusunda pazarlama verimliliğini arttırmak için makine öğrenimi algoritmalarını kullanmışlardır [7]. Çalışmada çevrimiçi içerik otomatik olarak etiketlenmesi sağlanmış; yapay sinir ağı yöntemi ile en yüksek doğruluğa %70 değeri ile ulaşılmıştır. Can 2021 yılında yaptığı çalışmada üniversite sınavına katılan katılımcıların katılımından oluşan anket sonucunda elde ettiği veri seti ile üniversiteye yerleştirilmeye hak kazanılmasının farklı makine öğrenmesi algoritmaları kullanarak sonuçları karşılaştırmıştır [8]. Çalışmada Lojistik Regresyon, Karar Ağaçları, Rastgele Orman, Destek Vektör Makineleri, K-En Yakın Komşu ve Gaussian Naif Bayes algoritmalarını kullanmıştır. Bu karşılaştırma sonucunda algoritmalar

arasında en yüksek tahmin yapan algoritma %73.77 doğruluk değeri ile Gaussian Naif Bayes algoritması olmuştur. Kaynar ve ark. 2017 yılındaki bir çalışmada, müşteri davranışlarının incelenerek ayrıştırılması mümkün olan müşterileri kategori halinde ayırarak algoritmalar arasında karşılaştırma yapılmıştır. Bu çalışmada Destek Vektör Makineleri, Yapay Sinir Ağları ve Naif Bayes algoritmaları kullanılmış; çalışma sonucunda %91.35 doğruluk değeri elde edilerek en iyi skor yapay sinir ağları yöntemi ile sağlanmıştır [9]. Işık ve ark.'nın 2021 yılında yapmış olduğu çalışma, teknoloji kabul modeli kullanarak çevrimiçi süpermarket alışveriş yapan tüketicilerin, tüketim tercihlerini etkileyen faktörleri açıklamaya çalışmaktadır. Çalışma kapsamında verilerin analizinde basit doğrusal regresyonu ve korelasyon katsayısı metotları uygulanmıştır [10]. Danışmaz'ın 2020 yılında yaptığı çalışmada, Türkiye'de yaşayan 200 kişiye yöneltilmiş anket ile Covid-19 pandemisi sonrası tüketicilerin çevrimiçi alışveriş tercihine etkisini incelemektedir. Covid-19 salgını öncesi kullanıcılar tarafından en çok tercih ettiği ürün grubu olan giyimin yerini gıdaya bıraktığını tespit etmiştir. Buna ek olarak bireylerin tüm toplumu etkileyen salgın esnasında gıda ürün grubu konusunda tedarik sorunları yaşayacağı korkusu etkili olmuş ve bu korku ile stok yapma amacıyla bu kategoriye yöneldiğini belirtmiştir [11]. Nizam ve ark. 2014 yılında makine öğrenmesi yöntemleri kullanılarak sosyal medya kullanan kullanıcılara ait duygu analizi çalışması yapmışlardır. Bu çalışmada, Naif Bayes, Rastgele Orman, Sıralı Minimal Optimizasyon, Karar Ağacı ve K-En Yakın Komşu sınıflandırma algoritmaları modellenmiş ve karşılaştırma yapılmıştır. Sıralı Minimal Optimizasyon yöntemi, %72.33 doğruluk oranıyla en iyi performans gösteren algoritma olmuştur [12]. Saydan 2008 yılında yaptığı çalışmada, çevrimiçi alışveriş deneyimine sahip tüketiciler ile bu deneyime sahip olmayan tüketicilerin demografik özelliklerini ayırtmak, tüketicilerin geleneksel alışveriş ve çevrimiçi alışverişlerde algılanan risk ve faydaları saptamak ve daha önce çevrimiçi alışveriş deneyimi bulunmayan tüketicilerin çevrimiçi alışveriş tercih etme olasılığını arttıracak durumları tespit etmeyi amaçlamaktadır [13]. Hendri ve ark. 2020 yılında internet tabanlı bir uygulama üzerinden hasta şikayetlerinden türetilen doktor seçiminin K-En Yakın Komşu ve Karar Ağacı C4.5 algoritmaları kullanılarak yapılmasına ilişkin bir çalışma gerçekleştirmiştir. En başarılı sonucu %100 doğruluk oranı ile K-En Yakın Komşu algoritmasının elde ettiği gözlemlenmiştir. Çalışmada kullanılan diğer algoritma olan Karar Ağacı C4.5 algoritması ise %80 oranında doğruluk elde etmiştir [14]. Saygılı 2014 yılında yaptığı çalışmada, kullanıcıları motive eden faydacı ve hazcı etmenlerin çevrimiçi satın alma niyeti üzerine etkisini incelemiştir. Çalışma kapsamında 481 katılımcıya anket uygulanmış ve anket sonucunda oluşturulan veri seti SPSS programı aracılığı ile Yapısal Eşitlik Modellemesi yöntemi ile incelenmiştir. Araştırma sonucunda kullanıcıları motive eden faydacı ve hazcı etmenlerin çevrimiçi satın alma niyeti üzerinde etkili olduğu tespit edilmiştir. Faydacı alışveriş etmenlerinin hazcı alışveriş etmenlere oranla etkisinin daha fazla olduğunu saptamıştır [15].

Benzer çalışmalardan farklı olarak bu çalışmada en sık kullanılan makine öğrenimi yöntemlerinin uygulanmasının yanında Lojistik Regresyon algoritması gauss dağılımı ile kullanılarak skorlama yapılmıştır. Genellikle performans ölçümü ve bankacılık sektöründe tercih edilen skorumanın bu alanda kullanılması, çalışmanın öne çıkan özelliklerindedir.

## 2. MATERYAL VE METOT (MATERIAL AND METHOD)

### 2.1. Veriseti (Dataset)

Çalışmanın anakütlesi Türkiye'de yaşayan 18 yaş üstü bireyler olarak belirlenmiş ve toplamda 394 kullanıcı anket için katılım sağlamıştır. Kullanıcılara toplamda 20'si demografik, 32'si alışveriş deneyimi ile ilgili olan 52 soru yöneltilmiştir. Anket, Google Forms üzerinden Aralık 2021-Ocak 2022 tarihleri arasında çevrimiçi olarak tamamlanmıştır [1]. Kullanıcılara toplam 53 soru sorulmuştur. İlk soru çevrimiçi market alışveriş deneyimi olup olmadığı ile ilgilidir. Bu soruya verilen cevap evet ise geriye kalan 52 sorunun cevabı alınmıştır. Anket sorularından bir kısmı aşağıdaki gibidir:

- Cinsiyetiniz nedir?
- Medeni durumunuz nedir?

- Kaç çocuğunuz var?
- Yaşınız?
- Öğrenim durumunuz?
- Aylık geliriniz nedir?
- Çalışma durumunuz nedir?
- İkamet ettiğiniz şehir hangisidir?
- Hangi sektörde çalışıyorsunuz.

Anket sorularına verilen her yanıtın matematiksel bir değeri olduğundan anket sonucunda bir skor elde edilmektedir. Anket çalışmasında kullanılan ölçeklerin güvenilirlik ve geçerlilik analizleri yapılmış olup, analize uygun olup olmadığını ortaya koyan değerlerden Kaiser-Mayer-Olkin (KMO) değeri 0.840 ve Bartlett testi sonucu anlamlılık değeri  $P=0.000$  ve Cronbach's Alpha ( $\alpha$ ) değeri 0.852 olarak hesaplanmıştır.

Modelleme esnasında eğitim ve test setlerinin belirlenmesi için K çapraz sorgulama yöntemi kullanılarak  $K=5$  ve  $K=10$  için çalışma yapılmıştır.  $K=10$  için veri seti 10 eşit parçaya bölünmüştür ve algoritmadan 10 defa ölçüm alınmıştır. Her ölçümde 1 parça test için kalan 9 parça ise eğitim için kullanılmıştır. Bu yöntem ile birlikte K değeri 10 olarak belirlenerek, %90 eğitim, %10 test seti ile doğruluk ölçütlerinde en iyi değerler elde edilmiştir.

Ayrıca uygulanan modelleme için parametre seçimi yapılırken, çevrimiçi market üzerinde kişisel bakım ürünlerini satın alan veya almayan müşterilerin tüm değişkenler ile ilişki aranmaktadır. Bu durum için Gini Endeksi ve Kanıtın Ağırlığı açıklayıcı yöntemleri kullanılmıştır.

Gini Endeksi, dağılımın tekdüze durumunu ortaya koyan bir ölçüttür. Gini endeksinin matematiksel olarak formülü denklem 1'deki gibidir. Denklemde m sınıf sayısını, N tercih durumu için örneklerin sayısını,  $S_i$  ilgili durumun i sınıfındaki örneklerin sayısını ifade etmektedir.

$$GINI_{Tercih} = 1 - \sum_{i=1}^m \left( \frac{S_i}{N_{Tercih}} \right)^2, \quad GINI_{Tercihsizlik} = 1 - \sum_{i=1}^m \left( \frac{S_i}{N_{Tercihsizlik}} \right)^2, \quad (1)$$

$$GINI = \frac{1}{n} (|N_{Tercih}| GINI_{Tercih} + |N_{Tercihsizlik}| GINI_{Tercihsizlik})$$

Bir niteliğe veya grup seviyesine ait bağlantılı riskin ölçülmesinde kullanılan Kanıtın Ağırlığı, bu çalışmada "tercih eden" veya "tercih etmeyen" olarak belirlenmiş bağımlı değişkenin değeri ile doğrudan bağlantılıdır. Kanıtın Ağırlığı, tüm grup aşamalarında tercih eden referansların tercih etmeyen referanslara olan oranına odaklıdır. Bir değişkene ait Kanıtın Ağırlığı değerinin hesaplanması denklem 2'de belirtilmiştir. Denklemde, "ln" doğal logaritmayı, "Tercih Dağılımı" belirli bir gruptaki tercihlerin yüzdesini, "Tercihsizlik Dağılımı" ise belirli bir gruptaki tercihsizlik yüzdesini vermektedir.

$$Kanıtın\ Ağırlığı = \ln \left( \frac{Tercih\ Dağılımı}{Tercihsizlik\ dağılımı} \right) \quad (2)$$

Skorkart her bir bağımlı değişkenin bağımsız değişkeni ne ölçüde açıkladığını ifade etmektedir. Her bir değişkenin sahip olduğu değer için bir skor ataması yapılır. Yapılan skor atamasından sonra tüm skorlar toplanarak bir skorkarne elde edilir. Skorkarnede elde edilen toplam puanın yine model tarafından belirlenen toplam puan gruplamasına göre konumu incelenir. Sürece ait akış Şekil 1'de verilmiştir.



Şekil 1. Lojistik regresyon ve skoreleme süreci.

Figure 1. Logistic regression and scoring process.

## 2.2. Lojistik Regresyon ve Skoreleme (Logistic Regression and Scoring)

Lojistik Regresyon, sisteme problem ile ilgili giriş değerlerine karşılık olması gereken sınıf bilgisinin verilmesi ile öğrenme gerçekleştirilir. Bu sürecin sonucunda elde edilen model genelleme yeteneği ile daha önce karşılaşmadığı ve sınıf bilgisi olmayan herhangi bir giriş değeri için anlamlı öngörüler yapabilmektedir. Modelleme yapılırken, olması gereken en yakın sonuca ulaşabilmek amacıyla en uygun fonksiyonu bulmayı hedeflemektedir. Hatanın sisteme öğrenme olarak geri dönüşü sebebiyle, denetimsiz öğrenmeye nazaran süreç çok daha hızlıdır [16].

Lojistik regresyon az değişken sayısı üzerinden bağımlı değişken ile bağımsız değişken arasındaki ilişkinin oluşturulması mantığına sahip olan yöntemdir. Lojistik regresyonu, Lineer regresyondan ayıran en önemli özellik değişkenin normal dağılım, sabit varyans ve lineer olma durumu gibi kriterlere ihtiyacı olmadan modellemenin sağlanabilmesidir [17]. Lojistik regresyon sapmalarına ait en küçük kareleri minimize etmek yerine, bir durumun muhtemel gerçekleşme durumunu maksimize eder. Lojistik Regresyon yöntemi, odds ve odds değerlerinin matematiksel olarak logaritmasını uygulamaktadır. Odds, lojistik regresyon yöntemi için bir durumun meydana gelme olasılığının meydana gelmeme olasılığına oranı olarak kullanılmaktadır. Lojistik regresyon açısından önemli bir kavram olan logit kavramı, odds oranının doğal logaritması alınarak bulunur [18]. Bir durumun gerçekleşme  $p(x)$  ve gerçekleşmeme durumunu  $1-p(x)$  ifade eden odds denklem 3 ile bulunmaktadır.

$$\text{Odds} = \frac{p(x)}{1-p(x)} \quad (3)$$

Modelleme sonrası hazırlanan skor değerinin güvenilirliği Receiver Operating Characteristic - Alıcı İşletim Karakteristiği eğrisi (ROC) ve Gini Katsayısı gibi yöntemler yardımı ile değerlendirilmiştir. ROC eğrisi ürünü tercih eden ve tercih etmeyen kullanıcıların birbirlerine karşı hareketlerine odaklanmaktadır. Kişisel bakım alanında alışveriş yapmayı tercih edenlerin kümülatif oranına göre tercih etmeyenlerin yüzdesini tahmin etmektedir. Eğri üzerindeki noktaların her biri, eşige ait skoru ifade etmektedir. Dikey eksen tercih etmeyenlerin oranı, yatay eksen ise tercih edenlerin oranı gösterilmektedir.

Gini katsayısının sifıra eşit olması kişisel bakım kategorisinden alışveriş yapmayı tercih edenlerin ve tercih etmeyenlerin ayrımının rastgele olarak yapıldığını göstermektedir. Değerin bir olması durumu ise tercih eden ve tercih etmeyen ayrıştırmasının tam olarak yapıldığını ifade etmektedir. Bu sayede, tercih eden ve tercih etmeyen kullanıcılar arasında farklılaşma başarısı ve belirlenen eşik puanının uygunluğu değerlendirilmektedir. Elde edilen değer 0 ise tercih eden ile tercih etmeyen kullanıcılar arasındaki ayrıştırmanın başarısız olduğu anlamına gelirken, 100 değeri ayrıştırmanın etkili olduğunu ifade etmektedir [19].

Gauss Dağılımı literatürde normal dağılım ve çan eğrisi olarak da adlandırılmaktadır. Gauss Dağılımını etkileyen iki parametre ortalama ve standart sapma değerleridir. Normal dağılım gösteren değişkenlerin daha yüksek doğrulukta tahmin edilmesine olanak sağlamaktadır. Gauss dağılımında ortalama, mod ve medyan gibi değerlerin birbirine eşitliği ya da yakınlığı söz konusu ise normal bir dağılımdan bahsedilmektedir [20].

Günlük yaşamda gözlenen olayların çoğunun uyumlu olduğu Gauss dağılımı sürekli olasılık dağılımları içerisinde bulunmaktadır. Bu şekilde olan bir dağılımda aşağıdaki özellikler sıralanabilir;

- Aritmetik ortalama, mod ve medyan birbirine eşittir.
- Modun medyana eşit olduğunu gösteren eğrinin tepe noktası aritmetik ortalamayı ifade etmektedir.
- Aritmetik ortalamanın sağında ve solunda bulunan eğriler simetriktir. Her bir alanın değeri 0.5'tir.
- Verilerin çoğunluğu ortalama değerin etrafında kümelenir [21].

### 2.3. Karar Ağaçları (Decision Trees)

En iyi sonucu tespit etmeye çalışırken birçok durum üzerinden ilerleyen karar ağaçlarında her bir durum karar ağacının dallarını oluşturur. Karar verme süreci ağacın en ucu olarak sayılan yaprak düğümüne kadar devam eder. Ağacın kökünden yaprak düğümüne kadar olan yol hedefin sınıflandırıldığı "kural" olarak adlandırılmaktadır. Karar ağaçlarında temel olarak kullanılan yapı "eğer-ise" şeklinde devam etmektedir. Yaygın olarak;

- Yeni bir verinin hangi sınıfın üyesi olabileceğinin tahmininde,
- Tıp alanında çeşitli vakaların Yüksek-Orta-Düşük olacak şekilde risk gruplarının ayrıştırılmasında,
- Eğer-sonra yapısı ile ileride gerçekleşebilecek olayların tahmin edilebilmesi mekanizma kurallarının oluşturulmasında,
- Alt gruplara ait olan bazı ilişkilerin belirlenmesinde,
- Parametrik modeller için bir veri seti içerisinde kullanılmak üzere faydalı olacakların belirlenmesinde kullanılmaktadır [22].

### 2.4. Rastgele Orman (Random Forest)

Rastgele Orman yöntemi, ağaç-tabanlı sınıflandırma metodudur. Orijinal veri seti üzerinden yeni bir eğitim seti oluşturan Rastgele Orman, bütün değişkenler arasında optimal bölünme sayesinde tüm düğümleri bölmemektedir. Her bir düğüm, gelişigüzel olarak seçimi yapılan tahmin ediciler arasından en başarılısını bularak bölünmektedir. Ardından, isteğe bağlı eleman seçimi kullanılarak bir ağaç geliştirilir. Bu prosedür, Rastgele Orman'ı hassasiyet ve doğruluk bakımından performansı yüksek bir yöntem olmasını sağlamaktadır. Rastgele Orman, aynı zamanda son derece hızlı ve fazla takılmaya karşı güçlüdür. Bunun yanında modelleyen kişinin ihtiyaç duyduğu kadar çok ağacın oluşturulması bu yöntem ile mümkündür. Rastgele Orman algoritmasının başlatılması için büyütülecek ağaç sayısı ve düğümü bölmek için kullanılacak değişkenlerin sayısını belirten parametrelerinin tanımlanması gerekmektedir [23].

### 2.5. K-En Yakın Komşu (K Nearest Neighbor)

Sınıflandırma problemlerinin çözümünde kullanılan K-En Yakın Komşu algoritması, sınıflandırılmamış verinin sınıflandırılmış veriler ile belirlenen uzaklık ölçütü kullanılarak mesafelerinin hesaplanması mantığını taşımaktadır. Elde edilen uzaklık değeri sınıflandırılmış veriler ile sınıflandırılmamış veri arasındaki benzerliği ortaya koymaktadır. Bu değerler arasında en yakın k adeti üzerinden sınıf kararı verilmektedir [24].

Makine öğrenmesi algoritmaları içinde çokça bilinen ve tercih edilen K-En Yakın Komşu algoritması, verinin sahip olduğu bir özelliğin kendisine en yakın olan başka bir özellik ile arasındaki uzaklık-yakınlık ilişkisine bağlı olarak sınıflandırma yapan bir algoritmadır [25].

### 2.6. Gradyan Arttırılmış Ağaçlar (Gradient Augmented Trees)

Hem regresyon hem de sınıflandırma problemlerinde kullanabilen Gradyan Arttırılmış Ağaçlar algoritması (GAA), veriler üzerinden çıkarımda bulunmaya ihtiyaç duymadığı için görsel takip

algoritmalarının trafik sistemleri ve öneri mekanizmalarının optimize edilmesinde sık tercih edilen güçlü bir ağaç modelidir. Çeşitli niteliklerdeki veri setleri ile çalışabilmesi, tek bir karar ağacı karşılaştırmasında daha performanslı öngörü gücüne sahip olması, daha büyük veri kümelerini işleme esnekliği ve dengesiz veri setlerindeki optimizasyon başarısı gibi özellikleri GAA algoritmasının güçlü yönleri olarak söylenebilmektedir. Yöntemin asıl amacı, çoklu yinelemeler yoluyla çok sayıda zayıf olan sınıflandırıcılar oluşturmasıdır. Bu sınıflandırıcılar her yinelemede güçlendirilmektedir. Elde edilen değerleri daha iyi bir sonuç haline getirebilmek için yeni oluşturulacak olan kombinasyon modeli gradyan yönünde oluşmaktadır. Bu işlem kalıtım azaltma işlemi olarak isimlendirilmektedir [26].

### 3. DENEYSEL ÇALIŞMALAR VE BULGULAR (EXPERIMENTAL STUDIES AND FINDINGS)

Çalışmada, Python programlama dili üzerinden Numpy, Matplotlib, Sklearn kütüphaneleri kullanılmıştır. Ön işlem adımında, veri setinde bulunan eksik veriler temizlenip ayrıştırılarak veriler modellemeye uygun hale getirilmiştir. Veri seti içerisinde KATEGORİ6 isimli değişken (Kişisel Bakım) tahmini için "1" değeri "Tercih Eden", "0" değeri için "Tercih Etmeyen" olarak etiketlenmiştir. Ardından K Çapraz Sorgulama yöntemi ile K=10 için veri setinde bulunan veriler %90 eğitim seti, %10 test seti olarak ayrıştırılmıştır. Veri seti uygulama için hazırlandıktan sonra çalışmada Lojistik Regresyon, Karar Ağacı, K-En Yakın Komşu, GAA ve Rastgele Orman yöntemleri ile modelleme yapılmıştır. Çalışmada skorkart uygulaması sadece Lojistik Regresyon modeli için modellenmiştir. Regresyon işleminin mantığı, problemin hangi durumların etkisinde olup olmadığının analizidir. Lojistik regresyon yöntemi, bir bağımlı değişkenin hem bağımsız değişken ile hem de bağımlı değişkenin odds değerinin logaritması ile doğrusal ilişkisi bulunması tezi ile meydana gelmektedir. Bağımsız değişkendeki değişimlerin, bağımlı değişkenin odds değerinin logaritmik sonucu üzerinde ne kadar farklılık göstereceğini lojistik regresyon yöntemi neticesinde elde edilen katsayılar ifade edeceğinden ötürü analiz zorlaşmaktadır. Bu durumdan ötürü elde edilen katsayılar için skor uygulaması ile dönüşüm işlemi yapılarak anlam çıkarabilme konusunda yaşanan zorluk hafifletilebilmektedir. Skor uygulaması ile skorkart lojistik regresyon yöntemi ile birlikte kullanıldığında modellemenin güvenilirliği ve uygulanabilirliği konusunda ciddi bir yol gösterici olmaktadır [19].

İlk olarak bağımsız değişkenler ile hedef değişkeni açıklayan ve değişkenlerin birbirleri arasındaki anlamlılık gücünü tespit etmek amacıyla Lojistik regresyon yöntemi uygulanmıştır. Modelleme neticesinde anlamlı değişkenler Çizelge 1'de verilmiştir.

Çizelge 1'de gösterilen  $Pr > Chisq$  ifadesi bir değişkenin diğer tüm değişkenler ile hedef değişkeni hangi oranda açıkladığını ifade etmektedir.  $Pr > Chisq$  sifıra yakınlığı değişkenin anlam değerinin fazlalığını göstermektedir. Çalışmada belirlenen alt sınır %95 olarak belirlenmiş olup, %95 altında anlamlı olan değişkenler regresyon sonuçlarına göre elenmiştir. Anlamlı bulunan değişkenlerin birbirleri arasındaki ilişkinin yönünü ve gücünü tespit etmek amacıyla korelasyon uygulanmış; ortaya çıkan korelasyon matrisi Çizelge 2'de gösterilmiştir. Çizelge 1, Çizelge 2 ve Çizelge 3 için kullanılan "AD2" alışveriş değerini, "AF6" algılanan faydayı, "AR2" algılanan riski, "INT\_GUN\_SURE" günlük internet kullanım süresini, "K\_KAPI\_ODEME" çevrimiçi alışverişte ödeme yöntemi seçimini, "YAS" yaşı ifade etmektedir.



**Çizelge 1.** Lojistik regresyon modeli sonucunda anlamlı bulunan değişkenler.*Table 1. Variables found to be significant as a result of the logistic regression model.*

Değişken	Chi-Square	Pr > ChiSq
AD2	116.756	0.0006
YAS	97.595	0.0018
K_KAPI_ODEME	74.703	0.0063
AR2	67.030	0.0096
INT_GUN_SURE	61.282	0.0133
AF6	52.196	0.0223

**Çizelge 2.** Korelasyon matrisi sonuçları*Table 2. Correlation matrix results.*

Parameter	AD2	AF6	AR2	INT_GUN_SURE	K_KAPI_ODEME	YAS
AD2	1	-0,18	-0,033	-0,043	0,029	0,069
AF6	-0,18	1	-0,03	0,008	0,099	0,053
AR2	-0,033	-0,03	1	0,065	0,08	0,026
INT_GUN_SURE	-0,043	0,008	0,065	1	0,178	0,013
K_KAPI_ODEME	0,029	0,099	0,08	0,178	1	-0,025
YAS	0,069	0,053	0,026	0,013	-0,025	1

Değişkenler arasındaki korelasyon incelendiğinde %80'in üzerinde değere sahip olan değişken bulunmadığı tespit edilmiş ve lojistik regresyon yöntemi sonucunda elde edilen bağımsız değişkenlerin tümü algoritmaya dahil edilmiştir.

Daha sonraki adımda ise skorkart dağılımı incelenmiştir. Modelin atamış olduğu skorlar ile oluşan skorkart kullanıcı dağılımı Şekil 2'de verilmiştir. Buna göre çoğunluğun orta değerde dağılması, en düşük ve en yüksek değerlerin en az sayıda kullanıcı ile sınırlı kalması modelin doğru kurulduğuna ve gauss dağılımını net şekilde yansıttığını göstermektedir.

Anlamlı bulunan değişkenler için, gruplara ayırma işlemi uygulanmıştır. Bu sonuca göre, kişisel bakım alışverişini tercih eden ve etmeyen dağılımlarına bakılarak skor ataması yapılmıştır (Çizelge 3).

**Şekil 2.** Skorkart kullanıcı adet dağılımı.*Figure 2. Distribution of scorecard users.*

**Çizelge 3.** Bağımlı değişken skorkart sonuçları.*Table 3. Dependent variable scorecard results.*

Değişken	Değişken Değeri	Skorkart Sonucu	Tercih Sonucu
AD2	AD2 < 3	43	29.81
	3 ≤ AD2 < 4	49	36.76
	4 ≤ AD2	60	49.48
AF6	AF6 < 2	7	10
	2 ≤ AF6 < 3	42	30.38
	3 ≤ AF6 < 4	45	33.33
	4 ≤ AF6 < 5	51	38.42
	5 ≤ AF6	71	57.69
AR2	AR2 < 3	75	59.38
	3 ≤ AR2 < 4	53	39.47
	4 ≤ AR2 < 5	49	36.55
	5 ≤ AR2	43	31.5
INT_GUN_SURE	INT_GUN_SURE < 2	44	32.74
	2 ≤ INT_GUN_SURE < 3	45	33.75
	3 ≤ INT_GUN_SURE	74	56.06
K_KAPI_ODEME	K_KAPI_ODEME < 2	57	42.05
	2 ≤ K_KAPI_ODEME < 4	50	37.59
	4 ≤ K_KAPI_ODEME	28	24.68
YAS	YAS < 6	53	39.64
	6 ≤ YAS < 7	49	36.84
	7 ≤ YAS < 8	26	21.05
	8 ≤ YAS	9	13.04

Skor dağılımı için her bir değişkenin değer gruplarından aldığı puanlar toplanmaktadır. Bu durumun neticesinde, kullanıcının tercih etme ya da tercih etmeme eğilimine göre nihai skor oluşturulmaktadır. Skor elde edilmesi ile kullanıcının kişisel bakım kategorisinden alışveriş yapma ya da alışveriş yapmama öngörüsüne göre gruplama yapılmıştır. Gruplar kullanıcıların hangi oranda ihtimalle alışveriş yapmayı tercih edeceğine dair ifadeleri içermektedir:

- Çok yüksek (Skor >= 338)
- Yüksek (304 ≤ Skor < 338)
- Orta (270 ≤ Skor < 304)
- Düşük (242 ≤ Skor < 270)
- Çok düşük (Skor < 242)

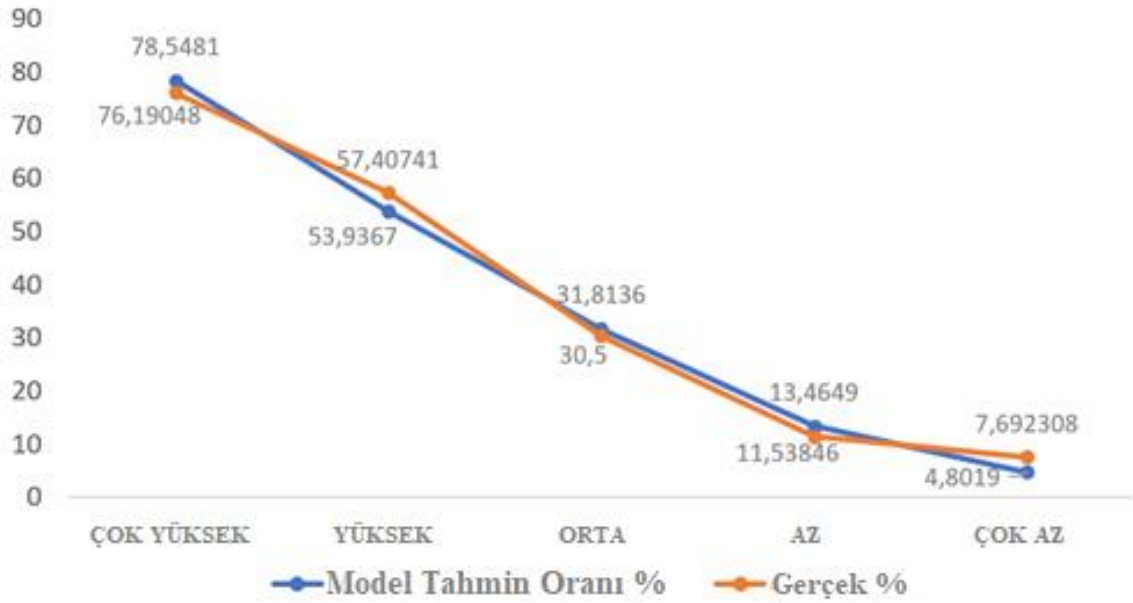
Skor puanları ve modelleme neticesinde bulunan skorkartın hesaplanması için regresyona ait parametrelerin ölçeklenmesi işlemi bu adımda yapılmıştır. Bu algoritma ile uygulanan modelin tahminine ilişkin sonuçlar Çizelge 4'te verilmiştir. Kullanıcıların skorkart sonucunda elde ettikleri toplam skor arttıkça satın alma ihtimali de artmaktadır.

Kullanıcılara ait kişisel bakım kategorisinden alışveriş yapmayı tercih etme durumunun gerçekleşme oranları ile modelin tahmin ettiği gerçekleşme durumu arasındaki ilişki Şekil 3'te verilmiştir.

Çizelge 4. Model tahmin sonuçları.

Table 4. Model prediction results.

Skor Adı	Skor Aralığı	Sayaç	Tercih Etme	Tercih Etmeme	Model Tercih Oranı %	Gerçek %
Çok Yüksek	Skor>=338	21	16	5	78,548	76,19
Yüksek	304<=Skor<338	108	62	46	53,937	57,407
Orta	270<=Skor<304	200	61	139	31,814	30,5
Düşük	242<=Skor<270	52	6	46	13,465	11,538
Çok Düşük	Skor<242	13	1	12	4,802	7,6923

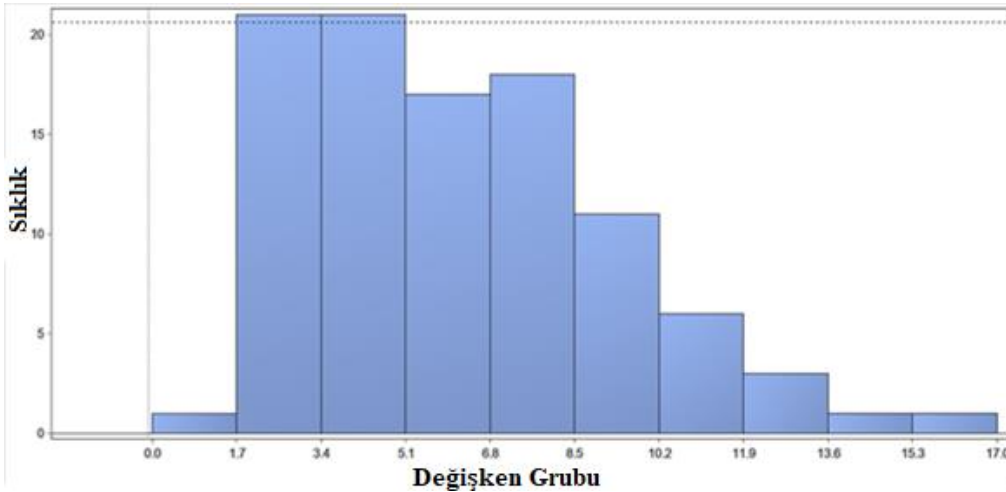


Şekil 3. Modelin tahmin ve gerçekleştirme durumu.

Figure 3. Model prediction and realization status.

Çalışmada daha sonraki adımda ise aynı problem Karar ağacı yöntemi ile modellenmiştir. Oluşturulan modelde ağacın 7 derinlik ve 56 yaprak sayısı bulunmaktadır. Değişkenlerin sıklık değerleri modeli açıklamak için yeterli düzeyde olduğu tespit edilmiştir. Modelde yer alan her bir değişken grubunun birden fazla benzersiz değerde cevap içerdiği Şekil 4'te gözlemlenmektedir.

Önemlilik değerleri yüksek olan değişkenler ve bölme kuralı sayı adetlerine ilişkin örnek Çizelge 5'te verilmiştir. Çizelge 6'da "KATEGORİ7" ev&yaşam kategorisini, "KATEGORİ8" ev&bakım&temizlik kategorisini, "KATEGORİ4" atıştırma kategorisini, "G4" alışverişteki gizlilik durumunu, "T3" sosyal tutumu, "SURE\_CEVIRIMICI\_ALISVERIS" çevrimiçi alışveriş kullanım geçmişini, "KATEGORİ5" süt&kahvaltı kategorisini ve "SEKTÖR" çalışılan sektör bilgisini ifade etmektedir.



Şekil 4. Karar ağacı sıklık grafiği.

Figure 4. Decision tree frequency graph.

Çizelge 5. Bölme kural sayı adetleri.

Table 5. Number of partition rules.

Değişken Adı	Bölme Kural Sayısı	Önem Değeri
KATEGORI7	17.0	1
KATEGORI8	12.0	0,8458
KATEGORI4	5.0	0,5665
G4	8.0	0,4577
YAS	8.0	0,4115
K_KAPI_ODEME	5.0	0,4112
T3	6.0	0,396
SURE_CEVIRIMICI_ALISVERIS	4.0	0,3771
KATEGORI5	2.0	0,3661
AR2	7.0	0,3603
SEKTOR	5.0	0,3584

Çalışmanın son adımında ise modelleme için K-En Yakın Komşu, GAA ve Rastgele Orman yöntemleri uygulanmıştır. Uygulanan makine öğrenmesi yöntemlerine ait parametreler Çizelge 6'da verilmektedir.

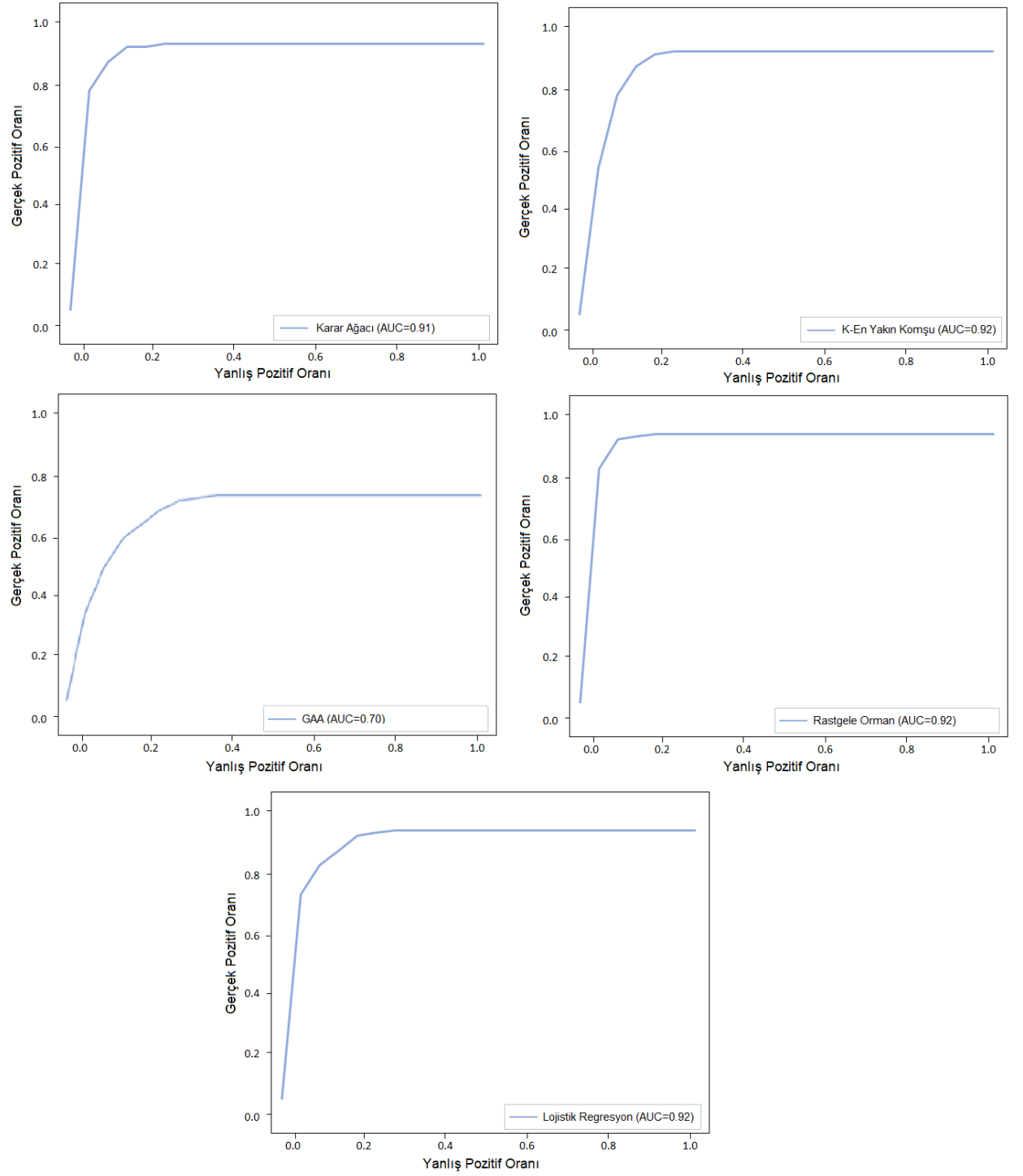
Çizelge 6. Uygulanan yöntemlerin parametre değerlerinin kullanımı.

Table 6. Use of the parameter values of applied methods.

Yöntem Adı	Parametrik Kullanım
Lojistik Regresyon	{'C': 2.7825594022071245, 'penalty': 'l2'}
Karar Ağacı	{'criterion': 'gini', 'max_depth': 40, 'max_features': 8, 'random_state': 50}
K-En Yakın Komşu	{'metric': 'euclidean', 'n_neighbors': 3, 'weights': 'distance'}
GAA	{'learning_rate': 0.03, 'max_depth': 8, 'n_estimators': 1000, 'subsample': 0.2}
Rastgele Orman	{'criterion': 'gini', 'max_depth': 8, 'max_features': 'auto', 'n_estimators': 200}

Çalışmada uygulanan tüm yöntemler ile ilgili elde edilen doğruluk ölçütleri, performans karşılaştırılması ortaya konmuştur. Uygulanan yöntemlerin doğruluk, AUC, Hassasiyet, F<sub>1</sub>, Geri Çağırma değerlerinin karşılaştırılması Çizelge 7'de gösterilmiştir.

Çalışmada uygulanan tüm yöntemlere ait ROC eğrisi grafikleri Şekil 5'te gösterilmiştir.



Şekil 5. Uygulanan yöntemleri ROC ve AUC değerleri.

Figure 5. ROC and AUC values for applied methods.

Çizelge 7. Uygulanan yöntemlerin performans karşılaştırılması.

Table 7. Performance comparison of the applied methods.

Yöntem Adı	Doğruluk	ROC-AUC	Hassasiyet	F <sub>1</sub>	Geri Çağırma
Lojistik Regresyon	0.831	0.90	0.829	0.84	0.832
Karar Ağacı	0.912	0.91	0.907	0.90	0.916
K-En Yakın Komşu	0.921	0.92	0.918	0.92	0.919
GAA	0.704	0.70	0.704	0.70	0.70
Rastgele Orman	0.928	0.92	0.921	0.92	0.924

#### 4. SONUÇLAR ve TARTIŞMALAR (RESULTS and DISCUSSIONS)

Literatürde, ilgili alanda benzer veri setleri ile genellikle tüketici tercihlerinin açıklanmasına yönelik çalışmalar yapılmıştır. Bu çalışmada, bir kullanıcının bir kategoriden alışveriş yapma olasılığını öngörülmesi ve bu tercihi etkileyen demografik özellikler üzerine odaklanılmıştır. Bu açıdan yaklaşıma sahip fazla sayıda çalışmalara rastlanmadığından, çalışmanın alanında özgün bir araştırma olduğu ifade edilebilmektedir. Ayrıca çalışmadaki diğer özgün katkı ise genellikle e-tedarik sistemlerinin performans ölçümü, bankacılık sektöründeki kredi risk yönetimi, gibi konularda kullanılmış olan skorkart ve skorkarne terimlerinin bu alanda kullanılmasıdır.

Çalışmada seçilen problemi modellemek için sık kullanılan makine öğrenmesi algoritmalarından Lojistik Regresyon, Karar Ağacı, K-En Yakın Komşu, GAA ve Rastgele Orman yöntemleri uygulanmıştır. Her yöntem, her bir değişkenin bağımlı değişken ile olan anlamlılığını ölçmektedir. Uygulanan tüm yöntemler için ilk olarak 52 değişken dikkate alınmış olup; anlamlılık durumuna göre analiz yapılmıştır. Karar Ağacı, K-En Yakın Komşu, GAA ve Rastgele Orman algoritmalarındaki sınıflandırma sonucunda en anlamlı sonucu veren 11 değişken belirlenmiştir. Belirlenen bu 11 değişkenin tamamının kendi içerisinde anlamlı olmadığı belirlenmiştir. Ancak lojistik regresyon ve skorkart uygulaması ile 6 değişkenin tamamının kendi içerisinde de anlamlı olduğu sonucu ortaya çıkmıştır. Sonuçların karşılaştırılması sadece tek bir skorkart uygulaması için yapılmıştır.

Modellere ait sonuçlar incelendiğinde Rastgele Orman yöntemi ile uygulanan modelinin 0.928 doğruluk oranı ile en yüksek performansı gösterdiği ve 0.921 AUC değeri ile modelin yüksek bir sınıflandırma başarısına ulaştığı görülmektedir. En düşük performansın ise 0.704 doğruluk oranı ve 0.702 AUC değeri ile GAA algoritmasına ait olduğu gözlenmiştir.

Ayrıca çalışmada Makine öğrenmesi algoritmalarından Lojistik regresyon ile gauss dağılımı uygulandığında 3 ana demografik özelliğin etken olduğu analiz edilmiştir. Bunun yanında çalışmada oluşturulan skorkart ile de elde edilen bu sonuçlar desteklenmektedir. Değişkenlerin, hedef değişken ile olan ilişkisi uygulanan yöntemler ile ortaya koyulmaya çalışılmıştır. Çalışmada lojistik regresyon yöntemi, skorkart ile modellendiğinde 6 değişkenin tamamının kendi arasında anlamlı ve ilişkili olduğu sonucu elde edilmiştir. Kullanıcıların demografik özelliklerinden olan yaş, günlük olarak internette geçirdikleri süre ve kapıda kredi kartı ile ödeme yapmayı tercih etme durumu kişisel bakım kategorisinden alışveriş yapmayı doğrudan etkileyen faktörler arasında yer aldığı elde edilen bulgular ile ortaya konmuştur.

Çalışmada elde edilen analiz bulgularına göre; çevrimiçi market alışverişinde kişisel bakım kategorisi için, 43 ile 47 yaş aralığından daha küçük yaşta olup, günlük internet kullanımı süresi 7-10 saat arasında olan ve ayrıca kapıda kredi kartı kullanım seçeneğini "Hiç kullanmam" veya "Nadiren kullanırım" şeklinde tercih eden kullanıcılar için kampanya uygulanması halinde hedef kitleye ulaşılmış olacaktır. Oraya konan bu ifadeler neticesinde, çalışmanın en büyük katkılarından biri kişisel bakım kategorisinden alışveriş yapabilecek kullanıcının sahip olması gereken özellikleri açıkça ortaya koyabilmektir.

Yapılan çalışmada perakende sektöründe faaliyet gösteren ve çevrimiçi market uygulaması ya da web hizmeti sunan tüm firmaların fayda sağlayabilmesi amaçlanmıştır. Buna göre elde edilen veriler ile

ilgili kategoriden yapılan satışların arttırılması ve bu kategori ile ilgili yapılacak olan kampanyaların başarıya ulaşması adına bu çalışmanın fayda sağlayacağı düşünülmektedir.

#### **Etik Standartlar Bildirimi (Declaration of Ethical Standards)**

Çalışma etik standartlara uygun olarak gerçekleştirilmiştir.

#### **Yazar Katkı Beyannamesi (Credit Authorship Contribution Statement)**

Tüm yazarlar usulüne uygun olarak makaleye katkıda bulunmuştur. Yazarlar yüksek lisans tez öğrencisi ve tez danışmanıdır. Burak Bahçivan: Veri elde etme, veri modeli oluşturma, deneylerin tasarlanması, araştırma, deneysel çalışmaların yürütülmesi, analiz Atınç Yılmaz: Veri modeli oluşturma, analiz, orijinal taslak yazma-düzenleme-revizyon, kontrol

#### **Çıkar Çatışması Beyannamesi (Declaration of Competing Interest)**

Yazarlar herhangi bir çıkar çatışması olmadığını beyan etmektedir.

#### **Destek / Teşekkür (Funding / Acknowledgements)**

Yazarlar, çalışmaya değerli zamanlarını ayırarak katkılarını sunan dergi editörleri ve hakemlere teşekkür etmektedir. Çalışma, Doç.Dr. Atınç Yılmaz'ın danışmanlığında yürütülen Burak Bahçivan'ın yüksek lisans tezinden türetilmiştir.

#### **Veri Kullanılabilirliği (Data Availability)**

Veriseti çevrimiçi market davranışları konusunda anket yolu ile Google Forms üzerinden Aralık 2021-Ocak 2022 tarihleri arasında elde edilmiştir; uygulanamaz.

#### **KAYNAKLAR (REFERENCES)**

- [1] B. Bahçivan, "Demografik Özelliklerin Online Market Kullanımına Etkisinin Makine Öğrenmesi Yöntemleri ile Tahmini", Yüksek Lisans Tezi", *Beykent Üniversitesi Lisansüstü Eğitim Enstitüsü*, İstanbul, 2022.
- [2] Y. Wang, J. Lin, B. Sheng, C. Sun, L. Si, and X. Liu, "Adaptive multi-task positive-unlabeled learning for joint prediction of multiple chronic diseases using online shopping behaviors", *Expert Systems with Applications*, Vol. 191, pp. 116232, 2022.
- [3] N. N. Moon, I. M. Talha, and I. Salehin, "An advanced intelligence system in customer online shopping behavior and satisfaction analysis", *Current Research in Behavioral Sciences*, Vol. 2, pp. 100051, 2021.
- [4] D. Koehn, S. Lessmann, and M. Schaal, "Predicting online shopping behaviour from clickstream data using deep learning", *Expert Systems with Applications*, Vol. 150, pp. 113342, 2020.
- [5] L. Wang, S. He, S. Su, Y. Li, L. Hu, and G. Li, "Urban neighborhood socioeconomic status (SES) inference: A machine learning approach based on semantic and sentimental analysis of online housing advertisements", *Habitat International*, Vol. 124, pp. 102572, 2022.
- [6] G. Volkmar, P. M. Fischer, and S. Reinecke, "Artificial Intelligence and machine learning: Exploring drivers, barriers, and future developments in marketing management", *Journal of Business Research*, Vol. 149, pp. 599-614, 2022.

- [7] J. Salminen, V. Yoganathan, J. Corporan, B. J. Jansen, and S. G. Jung, "Machine learning approach to auto-tagging online content for content marketing efficiency: A comparative analysis between methods and content type", *Journal of Business Research*, Vol. 101, pp. 203-217, 2019.
- [8] S. Can, "Lise öğrencilerinin üniversiteye giriş başarılarının eğitsel veri madenciliği ile tahmini", Yüksek Lisans Tezi, Beykent Üniversitesi Lisansüstü Eğitim Enstitüsü, 2021.
- [9] O. Kaynar, M. F. Tuna, Y. Görmez, and M. A. Deveci, "Makine öğrenmesi yöntemleriyle müşteri kaybı analizi", *Cumhuriyet Üniversitesi İktisadi ve İdari Bilimler Dergisi*, Vol. 18, No. 1, pp. 1-14, 2017.
- [10] P. Işık, and M. Öz, "Online alışveriş yapan süpermarket tüketicilerinin tüketim tercihlerinin teknoloji kabul modeliyle açıklanması", *OPUS Uluslararası Toplum Araştırmaları Dergisi*, Vol. 18, pp. 1538-1572, 2021.
- [11] A. T. Danışmaz, "Covid-19 salgınının tüketicilerin online alışveriş tercihine etkisi", *Social Sciences Research Journal*, Vol. 9, No. 2, pp. 83-90, 2020.
- [12] H. Nizam, and S. S. Akın, "Sosyal medyada makine öğrenmesi ile duygu analizinde dengeli ve dengesiz veri setlerinin performanslarının karşılaştırılması", XIX. Türkiye'de İnternet Konferansı, pp. 1-6, 2014.
- [13] R. Saydan, "Tüketicilerin online alışverişe yönelik risk ve fayda algılamaları", *Elektronik Sosyal Bilimler Dergisi*, Vol. 7, No. 23, pp. 386-402, 2008.
- [14] V. Hendri, C. Mawardi, and D. N. Santun, "Website based application of doctor selection classification derive from patient complaints using the C4.5 method and k-Nearest neighbor", *IOP Conference Series: Materials Science and Engineering*, Vol. 1007, pp. 012124, 2020.
- [15] M. Saygılı, "Faydacı ve hazcı alışveriş motivasyonlarının online satın alma niyeti üzerine etkisi", Yüksek Lisans Tezi, Sakarya Üniversitesi Sosyal Bilimler Enstitüsü, 2014.
- [16] E. S. Erdem, "Ses Sinyallerinde Duygu Tanıma ve Geri Erişim", Yüksek Lisans Tezi, Başkent Üniversitesi Fen Bilimleri Enstitüsü, Ankara, 2014.
- [17] S. Şenel, and B. Alatlı, "Lojistik Regresyon Analizinin Kullanıldığı Makaleler Üzerine Bir İnceleme", *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi*, Vol. 5, No. 1, pp. 35-52, 2014.
- [18] Ö. Çokluk, "Lojistik Regresyon Analizi: Kavram ve Uygulama", *Kuram ve Uygulamada Eğitim Bilimleri*, Vol. 10, No. 3, pp. 1359-1407, 2010.
- [19] B. Kasapoğlu, "Kredi Riskinin Hesaplanmasında Skorlama Yaklaşımı", Yüksek Lisans Tezi, İstanbul Teknik Üniversitesi Sosyal Bilimler Enstitüsü, İstanbul, 2009.
- [20] M. Akıllı, N. Yılmaz, and K. G. Akdeniz, "Study of the q-Gaussian distribution with the scale index and calculating entropy by normalized inner scalogram", *Physics Letters A*, Vol. 338, No. 11, pp. 1099-1104, 2019.
- [21] D. G. Altman, and J. M. Bland, "Statistics notes: The normal distribution", *BMJ*, Vol. 310, No. 6975, pp. 298, 1995.
- [22] G. G. Emel, and Ç. Taşkın, "Veri madenciliğinde karar ağaçları ve bir satış analizi uygulaması", *Eskişehir Osmangazi Üniversitesi Sosyal Bilimler Dergisi*, Vol. 6, No. 2, pp. 221-239, 2005.
- [23] Ö. Akar, and O. Güngör, "Rastgele orman algoritması kullanılarak çok bantlı görüntülerin sınıflandırılması", *Jeodezi ve Jeoinformasyon Dergisi*, Vol. 106, pp. 139-146, 2012.
- [24] M. A. Pala, M. E. Çimen, Ö. F. Boyraz, M. Z. Yıldız, and A. F. Boz, "Meme kanserinin teşhis edilmesinde karar ağacı ve KNN algoritmalarının karşılaştırmalı başarımları analizi", *Academic Perspective Procedia*, Vol. 2, No. 3, pp. 544-552, 2019.



- [25] D. Kılınç, E. Borandağ, F. Yücalar, V. Tunalı, M. Şimşek, and A. Özçift, "KNN algoritması ve R dili ile metin madenciliği kullanılarak bilimsel makale tasnifi", *Marmara Fen Bilimleri Dergisi*, Vol. 28, No. 3, pp. 89-94, 2016.
- [26] L. Yang, X. Zhang, S. Liang, Y. Yao, K. Jia, and A. Jia, "Estimating surface downward shortwave radiation over china based on the gradient boosting decision tree method", *Remote Sensing*, Vol. 10, No. 2, pp. 185, 2018.