İŞLETME ARAŞTIRMALARI DERGİSİ
JOURNAL OF BUSINESS RESEARCH-TURK
2023, 15(2), 953-969
https://doi.org/10.20491/isarder.2023.1629

Makine Öğrenmesi Tekniklerinin Bütçe Verimliliğine Uygulanması Üzerine Bir Çalışma

A Study on the Application of Machine Learning Techniques to Budget Efficiency

Göksel Kıvanç DEMİREL [a] **Ali ŞEN** [b]

[a] İstanbul Kültür Üniversitesi, Türkiye. kivanc@panelmedya.com
[b] İstanbul Kültür Üniversitesi, Türkiye. profdr.alisen@gmail.com

MAKALE BİLGİSİ

Anahtar Kelimeler:
Makine Öğrenmesi
Gradyan Artırma Makineleri
XGBoost
Pazarlama
Bütçe Verimliliği

Gönderilme Tarihi 21 Haziran 2022
Revizyon Tarihi 12 Haziran 2023
Kabul Tarihi 20 Haziran 2023

Makale Kategorisi:
Araştırma Makalesi

ÖZET

Amaç - Bu çalışma, pazarlama amacıyla işletmeye faydalı olabilecek müşteri kitlesini, yüksek miktardaki satış ve promosyon verilerinden faydalanarak seçmeyi ve puanlandırmayı amaçlamaktadır.

Yöntem - Bu amaç doğrultusunda, markalara ve perakendecilere müşteri verileri üzerinden hizmet sağlayan Dunnhumby şirketinin bilimsel amaçlarda kullanılmak üzere sunmuş olduğu ve haftalık kahvaltı ürünlerinden elde edilen "Breakfast at the FRAT" başlığı altında toplanılan satış bilgileri çalışmanın deney veri setini oluşturmuştur. Pazarlama bütçesini tüm müşterilerine harcamak yerine sadece potansiyel müşteri kitlesine harcamasına imkân tanıyan XGBoost algoritması kullanılarak pazarlamanın daha etkin ve verimli olabileceği müşterilerin belirlenmesine yönelik özgün bir model önerilmiştir.

Bulgular - Analizi yapılan veriler 2011 ile 2019 yılları arasında 156 haftalık bir süreyi kapsamaktadır. Özellik sayısının ve karmaşıklık durumunun minimuma indirgendiği çalışma kapsamında, model performansına ait ölçüt parametreleri yüksek başarı oranlarına sahiptir. Bu oranlar pazarlamada kullanılacak bütçenin uygun müşteri kitlesine harcanmasına yönelik oluşturulan model için kullanılan algoritmanın uygun olduğunu ortaya koymaktadır.

Tartışma - Büyük verilerin makine öğrenmesi teknikleri ile analiz edilmesi sonucu ortaya çıkan bulguların veri bilimine katkılar sunacağı ve çalışmada izlenilen yöntemin işletmelerin finansal açıdan tahmin ve öngörüler yapabilecekleri bir bütçe destek sisteminin altyapısını oluşturacağı düşünülmektedir. Gerçek dünya verilerinden elde edilen ve yapılan satışlar üzerinden birkaç özellik grubunun etkisi kullanılarak pazarlama için ayrılacak bütçenin verimliliğinin arttırılmasına yönelik bu çalışmada, en iyi tahminin yapıldığı sınıflandırma algoritmasının belirlenerek veri bilimine katkı sağlanması ve rehberlik etmesi mümkündür. Yapılan çalışmanın altyapısının daha da geliştirildiği bir modelin işletmeler tarafından kullanılarak iş dünyasına katkı sağlanma imkanı da vadır.

ARTICLE INFO

Keywords:
Machine Learning
Gradient Boosted Machines
XGBoost
Marketing
Budget Efficiency

Received 21 June 2022
Revised 11 June 2023
Accepted 20 June 2023

Article Classification:
Research Article

ABSTRACT

Purpose - This study aims to select and score the customer base that can be beneficial to the business for marketing purposes by using the large amount of sales and promotion data.

Design/methodology / approach - For this purpose, the sales information collected under the heading "Breakfast at the FRAT", which was provided by Dunnhumby Company, which provides services to brands and retailers through customer data, for scientific purposes and obtained from weekly breakfast products, constituted the experimental data set of the study. An original model has been proposed to identify customers for whom marketing can be more effective and efficient by using XGBoost algorithm, which allows the marketing budget to be spent only on the potential customer rather than all its customers.

Findings - The analyzed data cover a period of 156 weeks between 2011 and 2019. Within the scope of the study in which the number of features and complexity conditions are minimized, the criterion parameters of the model performance have high success rates. These rates reveal that the algorithm used for the model created to spend the budget to be used in marketing to the appropriate customer group is appropriate.

Discussion - It is thought that the findings resulting from the analysis of big data with machine learning techniques will contribute to data science and the method followed in the study will form the infrastructure of a budget support system where businesses can make financial estimates and predictions. It is possible to contribute and guide data science by determining the classification algorithm with the best estimate in this study, aimed at increasing the efficiency of the budget to be allocated for marketing by using the effect of a few feature groups over the sales obtained and made from real world data. There is also the possibility of contributing to the business world by using a model in which the infrastructure of the study is further developed by businesses.

Önerilen Atıf / Suggested Citation
Demirel, G. K., Şen, A. (2023). Makine Öğrenmesi Tekniklerinin Bütçe Verimliliğine Uygulanması Üzerine Bir Çalışma, *İşletme Araştırmaları Dergisi*, 15 (2), 953-969.

1. GİRİŞ

Makine öğrenmesi (machine learning - ML), büyük miktarda elde edilen verilerin zaman içinde gelişen öngörülebilir algoritmalar oluşturmasına olanak tanıyan gelişmiş bir yapay zeka (artificial intelligence - AI) türüdür. Ana odak noktası, büyük miktardaki verilere erişebilen ve bunları kendi başlarına öğrenmek için kullanabilen bilgisayar programlarının geliştirilmesidir. Gündelik yaşantımızın ayrılmaz bir parçası olan teknolojik ürünlerde her ne kadar farkında olmasak da makine öğrenmesi algoritmaları kullanılmaktadır. Müşterilerinin geçmişte izledikleri filmlere göre film önerileri sunan Netflix veya geçmişte satın alınan kitaplara göre müşterilerine yeni kitaplar öneren Amazon, arama motorunda bir ürün araştırması yapıldığında ilgili ürünle ilişkili reklamları internet ekranına anında getiren Google Adverse ve bir şarkı dinlenildiğinde benzer şarkıları listeleyebilen Youtube'un işleyiş mekanizması gündelik hayatımızdan makine öğrenmesine verilebilecek en güzel örnekler arasındadır. Bu konuda (MacKenzie, I., Meyer, C. ve Noble, S. How ., 2013; Yu, 2019:4-6). Bu platformlarda en ufak bir bilgi için arama yapıldığında, şarkı dinlenildiğinde ve hatta fotoğraf çekildiğinde, makine öğrenmesi platformların arama motorlarının bir parçası hâlini alarak, her etkileşiminden sürekli olarak öğrenerek ve gelişerek kişiye özel öneriler sunabilmektedir.

Bu çalışma,aynı bu kişiye özel öneriler bağlamından hareketle, işletmeye özel öneriler üzerine odaklanmıştır. Pazarlama odaklı bütçe oluşturmada, işletmeye faydalı olabilecek müşteri kitlesini, yüksek miktardaki satış ve promosyon verilerinden faydalanarak seçmeyi ve puanlandırmayı amaçlamaktadır. Bu çalışmada, gerçek dünya verilerinden elde edilen ve yapılan satışlar üzerinden birkaç özellik grubunun etkisi kullanılarak pazarlama için ayrılacak bütçenin verimliliğinin arttırılması amaçlanmaktadır.

Günümüzde petabayt boyutlarında satış, perakende ve promosyon verisi ile çalışıldığını düşünürsek (Bradlow vd., 2017, s.93), çok sayıda öngörücüyü modele dahil eden, onlarca özelliği etkili biçimde işleyebilen ve her bir özelliği modellemelerde kullanılmasına olanak sağlayan gradyan artırma makineleri (gradient boosting machine - GBM), rastgele ormanlar (random forests), yüzlerce hatta binlerce özelliği etkin bir şekilde işleyen ve her bir özelliğe modellemede kullanım fırsatı veren algoritmalar, satış ve tahmin modellemeleri oluşturmada geleneksel yöntemlerden daha umut verici görünmektedir. Antipov ve Pokryshevskaya tarafından yüksek boyutlu verilerle yapılan talep modelleme çalışmasında, gradyan artırma makineleri, rastgele ormanlar ve elastik ağların kullanıldığı üç sınıfın performası karşılaştırılmış ve gradyan artırma makinelerinin diğer algoritmalardan daha iyi bir performans gösterdiği belirlenmiştir (Antipov ve Pokryshevskaya, 2020). Yine Jang tarafından iyi bir Ar-Ge bütçesinin belirlenmesi üzerine yapılan farklı bir çalışmada, gradyan artırma makineleri diğer makine öğrenmesi algoritmalarından (AutoML, dağıtılmış rastgele ormanlar ve derin öğrenme) daha iyi bir performans sergilemiştir (Jang, 2019). Son yıllarda potansiyel tahmin oluşturma üzerine yapılan bazı çalışmalarda makine öğrenmesi algoritmaları arasından gradyan artırma makinelerinin öne çıktığını göstermektedir. Bu nedenle, çalışma kapsamında en iyi tahmin performansına sahip yöntemlerden biri olan GBM algoritması seçilmiştir. Buna ilaveten, çalışma kapsamında performans başarısı literatürdeki farklı araştırmalar ile ortaya konulan GBM algoritmasını diğer makine öğrenmesi algoritmalarıyla kıyaslamak yerine, oluşturulan tahmin modeli üzerinde bu algoritmanın nasıl bir performans gösterdiğine odaklanılmıştır.

Bu çalışma kapsamında, dünya çapında önde gelen bir perakende şirketinin müşteri satış verilerinden elde edilen büyük boyuttaki veriler üzerinden çok sayıdaki özelliğin etkisi esas alınarak pazarlamada kullanılacak olan bütçenin verimliliğinin artırılması amaçlanmıştır. Bu amaca ulaşmak için öncelikle genel bir yol haritası (çerçeve) belirlenmiş, daha sonra belirlenen bu çerçeve içerisinde Dunnhumby şirketi tarafından Aralık 2011 ile Ocak 2019 tarihleri arasında 156 haftalık bir süre için toplanılan ve açık kaynak olarak dışarıya sunulan müşterilere ait pazarlama verileri üzerinden daha öncesinde pazarlamaya ayrılan ve başarı elde edilmiş olan potansiyel kitle işaretlenerek GBM algoritması ile bir model oluşturulmuştur. Belirlenen çerçevenin model analizleri KNIME Analytics Platform 4.2.1 ile yapılmış ve bu modelde bir sonraki dönemin bütçesini ilgilendiren müşterilerin tümü girdi olarak verilmiştir. Modelden geçen veriler etiketlenerek puanlanmıştır. Böylece en yüksek puanı ve başarı etiketini alan potansiyel müşterilerin belirlenmesi ile pazarlamada kullanılacak olan bütçenin uygun müşteri grubuna harcanması hedeflenmiştir.

Deney veri setinde satılan ürünlerin müşterilere gösterilip gösterilmediği veya tanıtılıp tanıtılmadığı ile ilgili reklam (display) değişkeni modelin hedef değişkeni olarak belirlenmiştir. Bu değişken aracılığıyla yeni verilerin tahmini yapılarak puanlanmıştır. Deney veri setindeki özellikler, ileri özellik seçimi ve geri özellik ekleme yöntemleri kullanılarak belirlenmiştir. Deney veri setindeki verilerin %70'lik kısmı modelin

eğitiminde ve %30'luk kısmı modelin test edilmesinde kullanılmıştır. Modelin başarısı GBM algoritması ile test edilmiş ve model performansının ölçümünde sınıflandırma ölçütünü hesaplayabilen ve algoritmanın değerlendirilmesi için önem arz eden karmaşıklık matrisi yönteminden yararlanılmıştır. Son olarak, GBM algoritması için model performansı ölçümüne ait ölçüt hesaplamaları veya sınıflandırma raporu elde edilmiştir.

Çalışmanın diğer bölümleri şu şekilde kurgulanmıştır. Çalışmanın amacının anlatılması, kavramsal çerçeve bölümünde makine öğrenmesinin tanımı ve genel özellikleri ile ilgili bilgilere yer verilmiş, daha sonra makine öğrenmesi algoritmaları ve çalışmanın amaçları doğrultusunda kullanılan gradyan artırma makineleri anlatılmıştır. Veri setine, özelliklerin seçimine ve sistemin akış mimarisine ilişkin ayrıntılı bilgiler yöntem bölümünde sunulmuştur. Verilerin analiz edilmesi sonucu elde edilen sonuçlara ait bilgiler bulgular kısmında belirtilmiştir. Çalışmanın alana sağladığı katkı ve güncel durum tartışma alanında, genel bir değerlendirmesi ve önerilerine ilişkin görüşlere de sonuç ve öneriler bölümünde yer verilmiştir.

2.KAVRAMSAL ÇERÇEVE

2.1 Makine Öğrenmesi ve Veri Kullanımı

Günümüzde makine öğrenmesinin tam potansiyelini veya ne kadar ilerleyebileceğini şu anda görmekten çok uzak olsak da, günden güne daha fazla büyüyen veri kümelerinin yakın bir gelecekte daha ileri modellemeleri tespit etmemize ve daha doğru bir şekilde sonuçları tahmin etmemize yardımcı olacağı öngörülmektedir. Makineler, çeşitli verilerdeki kavramları ve temaları kolayca belirleyebilir, duyguları ve insan iletişimlerini yorumlayabilir ve kullanıcılara yeterli yanıtlar oluşturabilir. Müşterilerin davranışlarını ve kararlarını kolayca tahmin ederek bu verileri gelecekteki sorunları çözmek için kullanabilir. Önümüzdeki yıllarda, pazarlama açısından daha akıllı aramalar, daha akıllı reklamlar, iyileştirilmiş içerik sunumları sağlama, sürekli öğrenme, botlara güvenme, dolandırıcılığı ve veri ihlallerini önleme, duygu analizi, görüntü ve ses tanıma, satış tahmini, dil tanıma, öngörülü müşteri hizmetleri, müşteri segmentasyonu gibi konularda daha büyük bir yapay zeka etkisi beklenebilir (Cuirieska S., Stankovska A. ve Efremova T, 2018:302-303).

Geride bıraktığımız son on yılda, ağa bağlı sistemleri kullanan teknolojik cihazların büyük miktarda veriler toplayabilmiş ve bu verilerin taşınmasında önemli ölçüde bir ilerleme kaydedilmiştir (Yang ve Zhang, 2018, s. 3181). Günden güne sürekli olarak büyüme gösteren bu verileri toplama çabasında olan araştırmacılar, verileri işleyerek veya analiz ederek faydalı öngörülere, tahminlere ve kararlara ulaşmak için makine öğrenmesi algoritmalarına yönelmişlerdir (Sun, Z.-L., Choi, T.-M., Au, K.-F. ve Yu., Y. (2008). Sales forecasting using extreme learning machine with applications in fashion retailing. decision support systems, 46, 411-419.). Makine öğrenmesinin pazarlama alanına uygulanması sayesinde hedef müşteri kitlesinin bir sonraki satın alma kararlarının takip edilmesi, tahminlerde bulunulması, öngörüler ile hızlı kararlar alınması, uygun stratejilerin belirlenmesi ve doğru bütçe planlarının oluşturulması gibi önemli katkıları bulunmaktadır (Cui, D. ve Curry, D. 2005). Makine öğrenmesi sayesinde müşterilerin satın almaları takip eden ve gerekli analizleri yapan işletmeler, müşteri eğilimlerini tahmin edebilmenin yanı sıra, bir sonraki müşteri davranışını da tahmin edebilir (Ma vd., 2016:256; Ma ve Fildes, 2017: 680). Makine öğrenmesinin kullanımı ile sorunlar karşısında güçlü çözümler üretmek, makinelerin dünyayı gerçekten insanlarla aynı şekilde anladığı bir çağda yaşadığımız anlamını taşır.

2.2 Makine Öğrenmesi ve Tahmin Modelleri

Makine öğrenmesi, tahmin modellerinin iyileştirilmesinde ve yönetimsel kararların alınmasında yarar sağlayabilecek yenilikçi bir yöntem olarak dikkat çekmektedir. Toplanılan verilerden tahmine yönelik doğru modellerin oluşturulmasını esas alan doğrudan pazarlama, makine öğrenmesinin öncelikli olarak kullanılabileceği bir alandır. Doğrudan pazarlamanın birçok işletme tarafından bir dağıtım stratejisi şeklinde benimsemiş olması, bu alandaki harcamaların artmasına neden olmuştur. Bu bağlamda, pazarlamacılar açısından satışların artırıldığı, maliyetlerin düşürüldüğü ve kârlılığın artırıldığı bir müşteri yanıt modelinin oluşturulması öncelikli bir ihtiyaç hâlini almıştır (Cui vd., 2006: 597). Araştırmacı veya analistler, müşterilerin satın alma davranışlarının tahmini ile ilgili geleneksel yaklaşımların yanında, büyük verilerin analiz edilmesinde çok sayıda avantajı içerisinde barındıran makine öğrenmesi tekniklerinden yararlanmaktadır. Talebin tahmin edilmesi ve satış etkenlerinin mantığının anlaşılması, perakende analitiğinin çözüm arayan problemleridir. Satışlardaki fiyat duyarlılığının yanında promosyonların gerçek etkisinin ne olduğunun bilinmesi, satış ilişkilerinin karmaşık olması sebebiyle cevabını arayan zor bir soru olarak güncelliğini korumaktadır. Tanıtımı

yapılmayan ürünlere tanıtımı yapılan ürünlerin engel oluşturması, promosyonların etkinliklerinin büyük ihtimalle daha önceki haftalarda yapılan promosyon faaliyetlerini etkilemesi, bazı ürünlere uygulanan promosyonların diğer ürün satışlarının üzerinde sinerjik bir etki oluşturması şeklindeki durumlar satış ilişkilerinin karmaşıklığı nedeniyle ortaya çıkabilecek olasılıklardır. Bir ürüne uygulanan promosyonun satışları artırması üzerine ya hiçbir ya da bir kısım etkisi perakendeciler açısından anlamlıdır (Gedenk, 2018:345). Bununla ilgili olarak %50'den fazla uygulanan tüm promosyonların perakendeciler açısından kâr ifade etmediği belirtilmiştir (Ailawadi vd., 2007: 566). Üreticiler tarafından promosyon uygulanması faydalı olmasına karşın, perakendeci tarafından negatif bir etkisinin ortaya çıkma olasılığı yüksektir. Zira promosyon uygulanmamış bir ürün, kârlılığı daha düşük olan promosyonlu bir ürüne dönüşür. Promosyonların etkili bir şekilde yönetilmesinde satışlar üzerine nasıl bir etkilerinin olduğu durumlarının ortaya çıkarılması gereklidir. Bu etkinin ortaya konulabildiği bir satış modeli tam olarak perakendecilerin ihtiyaçlarını karşılayabilecektir. Diğer taraftan, pazarlama araştırmalarında elde edilen büyük veride potansiyel olarak çok sayıdaki değişkenin birbiriyle ilişkili olan etkileşimlerinin anlaşılması gerekmektedir. İşletmeler üretim, pazarlama, envanter, finans ve bütçeleme gibi birçok farklı alanda iş kararları verirken büyük ölçüde stok tutma birimleri (stock keeping unit - SKU) düzeyinde doğru satış tahminlerinden yararlanmaktadır. Herhangi bir SKU'daki satış ve promosyon etkileri, çok sayıda başka kategorideki pazarlama ve satış faaliyetlerinden potansiyel olarak etkilenir. Bu bağlamda, hedef değişkenleri etkileyebilecek kategori içi ve kategoriler arasında çeşitli değişkenler arasından önemli değişkenleri büyük bir veri setininin içerisinden belirlemek ciddi bir modelleme zorluğunu ortaya çıkarmaktadır.

Talebin tahmin edilmesi ve satış etkenlerinin anlaşılması, pazarlamanın ve pazarlamayı yakından ilgilendiren bütçe planlamalarının en önemli işlevlerinden biridir. Daha az sayıdaki öngörücülü geleneksel modellerin satış modellemelerinde çoğunlukla kullanılmasına karşın, büyük verilerin kullanıldığı bütçe verimliliğine veya planlamasına yönelik sınırlı sayıda araştırma yapılmıştır (Ma, S., Fildes, R. ve Huang, T. (2016); Ma ve Fildes, 2017). Talep tahmininde büyük miktardaki verileri kullanmanın önemine vurgu yapan Ma ve arkadaşlarının yapmış oldukları çalışmada, özellik seçiminde dört adımlı bir yaklaşım yöntemi önerilmiş, ürünler ile dönemlerin arasındaki etkilerin hesaplanmasının önemine vurgu yapmışlardır. Bir sonraki adımda, bu yaklaşımlarını birçok dönemde kârlılığı artıracak kategorileri temel alan bir optimizasyon modelini ortaya koymuşlardır (Ma ve Fildes, 2017). Bu bağlamda, müşteri esaslı büyük boyutlu satış ve pazarlama verileri üzerinde birbiriyle ilişki içerisinde olan çok sayıda değişken veya özelliğin hesaba katıldığı, gerçek dünyadaki pazarlama problemleri üzerinden talebin tahmin edildiği ve pazarlamada kullanılacak bütçenin doğru müşteri kitlesine harcanması imkânının sağlandığı bir modelin oluşturulmasına gereksinim duyulmaktadır.

2.3. Makine Öğrenmesinin Pazarlamadaki Rolü

Makine öğrenmesi, deneyimler yoluyla otomatik olarak gelişen bilgisayarlar oluşturma yollarını ele alan bir yapay zeka alt kümesidir (Jordan ve Mitchell, 2015:255). Başka bir deyişle, makine öğrenmesi verilerdeki temel kalıpları öğrenmek ve bu kalıplara dayalı tahminlerde bulunmak için tasarlanmış yöntem veya algoritmaların incelenmesini ifade eder (Dzyabura ve Yoganarasimhan, 2018:255). Günümüzde büyük miktarlarda elde edilebilen verilerden yararlı bilgileri ortaya çıkarmak ve analize dayalı algoritmalar geliştirmek için bu verileri işleyebilmek çok önemlidir. Bu anlamda, makine öğrenmesinin veri eğilimlerine ve veriler arasındaki geçmiş ilişkilere dayalı algoritmalar tasarlamak için kullanılan yapay zekanın ayrılmaz bir parçası olduğu düşünülebilir. Makine öğrenmesi, insan öğrenimini taklit etmek için bilgisayarları kullanır ve bilgisayarların gerçek dünyadan bilgi tanımasına ve edinmesine ve bu yeni bilgiye dayalı olarak bazı görevlerdeki performansı artırmasına olanak tanır.

Yapay zekanın bir alt dalı olan makine öğrenmesi, bilgisayarlara adım adım problem çözdürücü veriler vermek yerine, onların farklı algoritmalar ile sonuç için ne yapacaklarını öğrenmelerini sağlamaktır. Bu durum belirlenmiş bir yolu takip eden standart programlar yerine, dinamik olarak ortaya çıkabilecek olası bir soruna daha önce tanımlanmış veya öğrenilmiş çok fazla sayıda içerikleri kullanarak çözüm bulan algoritmaların geliştirilmesi demektir (Gürsakal, 2017).Yapay zekanın bir parçası olan makine öğrenmesi ile ilgili ilk kavramlar 1950'lerde ortaya çıkmış ve birkaç basit algoritmadan başlayarak verilerden öğrenebilen makinelere sahip olma ilgisiyle makine öğrenmesi serüveni başlamıştır. 1960'lı ve 70'li yıllarda desen sınıflandırmasına odaklanan çalışmalar ile bu serüven devam etmiş ve insan zekasını taklit edebilen makinelerin ortaya çıkmasıyla klasik yapay zeka uygulamalarından ayrılmıştır. 1990'larda veriye dayalı bir yaklaşımla

pratik sorunları çözmenin bir yolu olarak popülerlik kazanmış, ayrı bir araştırma alanı olarak potansiyelini artırmış ve araştırmacıların daha gelişmiş algoritmaları ortaya koyabilmeleri için teşvik ederek her geçen gün daha da önem kazanarak günümüze gelmiştir. Bugün makine öğrenmesi algoritmaları; bilgisayar bilimlerinin yanı sıra istatistik, matematik, biyoenformatik, izinsiz giriş algılama, bilgi alma, oyun oynama, kötü amaçlı yazılım algılama, işletme, pazarlama, reklamcılık, biyoloji, tıp, uzay bilimleri ve sosyal bilimler de dahil olmak üzere birçok alanda kullanılmaktadır.

Makine öğrenmesinin pazarlama alanına ilişkin uygulamaları genel olarak talebin anlaşılması ve tahmine yönelik çalışmaları içermektedir (Brei, 2020:201). Tam ve Kiang'ın banka başarısızlığını tahmin etmeye yönelik yapmış olduğu çalışmada doğrusal sınıflandırıcı, lojistik regresyon ve k-en yakın komşuluk (KNN) algoritması gibi geleneksel yöntemlerden daha iyi performans gösteren bir yapay sinir ağı (YSA) yaklaşımı geliştirilmiştir (Tam ve Kiang, 1992). Cui ve Curry'nin yaptıkları çalışmada destek vektör makinesi (DVM) algoritmasının otomatik modelleme, toplu üretim modelleri, akıllı yazılım aracıları ve veri madenciliği gibi pazarlamada ortaya çıkan ortamlardaki sonuçları tahmin etme yeteneği araştırılmıştır. DVM'nin tahmin isabet oranlarının farklılıkları koşullu lojit modeli (KLM) ile karşılaştırılmıştır (Cui ve Curry, 2005). Chandukala ve arkadaşları, birleşik analiz bağlamında karşılanmayan talebi belirlemek ve zayıf tercihleri güçlü kategori ilgisinden ve heterojen beğeni etkilerinden ayırmak için bir model önermişlerdir (Chandukala vd., 2011). Heterojen bir değişken bölüm modeli algoritması izlenerek ve veriler arttırılarak modelde Naive Bayes tahmini gerçekleştirilmiştir. Jacobs ve arkadaşları, görev verilerini sıralamak için esnek bir çerçevenin yanı sıra, tüketicilerin bir dizi öğe arasında yaptığı çağrışımları keşfetmenin bir yolunu sağlayan özet yığınlarını belirlemek için yeni bir optimizasyon yaklaşımını tanıtmışlardır (Jacobs vd., 2016). Bu çalışmada, Monte Carlo simülasyonları ve ampirik bir uygulama kullanılarak, elde edilen prosedürün ölçeklenebilir hesaplama açısından verimli olduğu gösterilmiştir. Grushka-Cockayne ve arkadaşları, aşırı uyum ve aşırı güvenin ortalama tahmin üzerindeki birleşik etkisini incelemek için teorik bir model tanıtmışlardır (Grushka-Cockayne, 2017). Yapılan çalışmada fazla uyumlu ve aşırı güvenli yüzlerce regresyon ağacını bir araya toplamak için rastgele orman algoritması, olasılıkları kırpmak için ideal bir ortam olarak belirlenmiştir. Bilinen birkaç veri setini kullanarak, kırpılmış toplulukların rastgele orman algoritmasının tahmin doğruluğunu önemli ölçüde arttırdığı belirlenmiştir. Kaynar ve arkadaşları, telekomünikasyon sektöründeki müşteri kayıplarının tahmine ilişkin DVM, YSA ve Naive Bayes algoritmaları ile analizler gerçekleştirmiş ve bu analizler sonucunda YSA algoritmasının daha başarılı olduğu belirlenmiştir (Kaynar vd., 2017). Onan tarafından yapılan çalışmada, Twitter'in Türkçe dilindeki mesajların duygu analizleri DVM, YSA ve lojistik regresyon algoritmaları aracılığıyla yapılmıştır (Onan, 2017). Çalışmadaki en yüksek başarımın Naive Bayes algoritmasına ait olduğu tespit edilmiştir. Bulut tarafından bankacılık sektöründeki müşterilerin davranışsal ve yaşamsal özelliklerinin dikkate alındığı teorik bir çalışma yapılmıştır (Bulut, 2019). Teorik bir model oluşturmak için veri toplama yöntemleri, toplanılan verilerin sayısallaştırma durumu ve yöntemin nasıl uygulanacağına ilişkin bir yapay destek modeli önerilmiştir. Tuna ve arkadaşları, otel müşterilerinin yapmiş oldukları geri bildirimlerden faydalanarak müşteri duygusunun otelden alınan hizmetin derecelendirilmesi ne kadar uyumlu olduğuna ilişkin bir çalışma yapmışlardır (Tuna, M. F., Kaynar, O. Ve Akdoğan, M. Ş. 2021). Denetimli duygu sınıflandırmasına yönelik DVM, YSA, KNN, karar ağacı, rastgele orman, doğrusal diskriminant analizi ve lojistik regresyon algoritmaları kullanılmış ve en başarılı sonucu lojistik regresyon algoritmasının verdiği belirlenmiştir.

2.4 Makine Öğrenmesi Algoritmaları

Makine öğrenmesi algoritmaları, giriş değerleri ve sonuç çıktıları arasında bir ilişki belirlemek için kullanılır. Bu algoritmalar, model parametrelerini optimize etmeye çalışır. Günümüzde petabaytlarca verinin kolayca toplanması ve depolanmasının bir neticesi olarak veri biliminde makine öğrenmesine verilen önem seviyesi artmıştır. Buna ilaveten, toplanılan verilerdeki bilgilerin geniş kapsamlı olması, manuel olarak bu verilerin kontrolünü ve analizini olanaksız bir duruma getirmektedir. Bu açıdan bakıldığında, makine öğrenmesi algoritmalarının verilerin analiz edilmesinde büyük bir rolü vardır. Diğer taraftan, algoritmaların hesaplama maliyetini düşürmesi makine öğrenmesini popüler kılmaktadır. Makine öğrenmesi algoritmaları pazarlama alanında maliyet ve zaman bakımından avantaj sağlamaktadır. Bir makine öğrenmesi algoritmasına veya tekniğine karar vermek, gerçek yaşamdan toplanılan verileri kullanmak ve girdilerin özellik kalitesi algoritmalardan doğru ve istenilen sonuçların alınmasında önemlidir.

Makine öğrenmesi doğası gereği disiplinler arasıdır ve bilgisayar bilimi, istatistik, matematik, yapay zeka gibi çeşitli alanlardan teknikleri içerir. Makine öğrenmesinin temel özelliği, bilgisayar görüşü, yapay zeka ve veri madenciliğine dayanan algoritmalar uygulayarak deneyim verisi elde etmektir. İşlemci hızı ve bellek boyutunun artmasıyla son zamanlarda oldukça popüler hale gelen makine öğrenmesi artık öğrenmek, sonuç çıkarmak veya verileri çıkarmak için matematiksel veya istatistiksel analiz kullanan çok sayıda algoritmaya sahiptir. Yeni varyasyonların veya kombinasyonların önerildiği makine öğrenmesi algoritma sayısı ise her geçen gün artmaya devam etmektedir.

Bu çalışmada kullanılan makine öğrenmesi algoritması aracılığıyla pazarlamada kullanılacak olan bütçe verimliliğinin uygun müşteri grubuna harcanabilmesi için satış ve promosyon verilerinden analiz yapılarak tahminde bulunulmaktadır. Bu nedenle, makale kapsamında literatürde yer alan çok sayıda algoritmayı incelemek yerine, yalnızca verilerin analiz edildiği GBM algoritmasına odaklanılmıştır.

2.5 Gradyan Artırma Makineleri

Karar ağaçlarını temel öğrenenler olarak kullanan popüler topluluk yöntemlerinde biri gradyan artırma makineleri veya GBM'dir (Friedman, 2001, s. 1189). GBM, regresyon veya sınıflandırma ağacı modellerinden oluşan bir topluluktur. Her ikisi de, kademeli olarak iyileştirilen tahminler yoluyla tahmine dayalı sonuçlar elde eden ileriye yönelik toplu öğrenme yöntemleridir. GBM her ağacın bir öncekini öğrenip gelişmesiyle sığ ve zayıf ardışık ağaçlardan oluşan bir topluluk oluşturur. Bu zayıf ardışık ağaçlar birleştirildiğinde, genellikle diğer algoritmalarla yenilmesi zor olan güçlü bir "topluluk" oluşturur. Güçlendirme, ağaçların doğruluğunu artırmaya yardımcı olan esnek, doğrusal olmayan bir regresyon prosedürüdür. Artırımlı olarak değiştirilen verilere sırayla zayıf sınıflandırma algoritmaları uygulayarak, zayıf tahmin modellerinden oluşan bir topluluk üreten bir dizi karar ağacı oluşturulur. Ağaçları güçlendirmek doğruluklarını arttırırken, aynı zamanda hızı ve insan tarafından yorumlanabilirliği de azaltır. Gradyan artırma makineleri, bu sorunları en aza indirmek için ağaç artırmayı genelleştirir.

GBM, zayıf öğrenenleri güçlü öğrenenlere dönüştürme yöntemidir. Artırmada, her yeni ağaç, orijinal veri setinin değiştirilmiş bir versiyonuna uygundur. GBM'ye kolay bir şekilde açıklayabilmek için AdaBoost algoritmasının işlevi örnek olarak verilebilir. AdaBoost algoritması, her bir verinin eşit ağırlıkta olduğu bir karar ağacını eğiterek başlar. İlk ağacı değerlendirdikten sonra, sınıflandırması zor olan verilerin ağırlıklarını artırıp, sınıflandırması kolay olanların ağırlıklarını düşürür. İkinci ağaç bu ağırlıktaki veriler üzerinde büyütülür. Buradaki temel fikir, ilk ağacın tahminlerini geliştirmektir. Dolayısıyla yeni model Ağaç 1 + Ağaç 2 olur. Daha sonra, bu yeni iki ağaçlı topluluk modelinden sınıflandırma hatası hesaplanır ve revize edilmiş kalıntıları tahmin etmek için üçüncü bir ağaç yetiştirilir. Bu işlem belirli sayıda yineleme için tekrarlanır. Sonraki ağaçlar, önceki ağaçlar tarafından iyi sınıflandırılmamış verileri sınıflandırmaya yardımcı olur. Bu nedenle, nihai topluluk modelinin tahminleri, önceki ağaç modelleri tarafından yapılan tahminlerin ağırlıklı toplamı kadardır (Singh, 2018).

GBM; yaygın şekilde kullanılan, en iyi tahmin performansına sahip yöntemler arasında yer alan, birçok alanda başarısı kanıtlanmış ve dünyaca ünlü Kaggle yarışmalarını kazanmak için kullanılan son derece popüler bir makine öğrenmesi algoritmasıdır (Chen ve Guestrin, 2016:785). Çalışma kapsamında, gradyan artırma makinelerinin çok yönlü bir uygulaması olan aşırı gradyan artırma ağaçları (XGBoost) algoritması kullanılmıştır. XGBoost algoritmasının kullanılma nedenlerinden biri, büyük ölçekli verilerle uğraşmak için gelişmiş tekniklerin özel kullanımı nedeniyle çok iyi ölçeklenebilmesidir. XGBoost, başlangıçta regresyon problemleriyle başa çıkmak için tasarlanmıştır, ancak farklı amaç ve hedef işlevlerini ve sayısal tahminlerin yorumunu uygun şekilde uyarlayarak tanımlayabilir. Her model önceden tanımlanmış sayıda karar ağacından oluşur. Bu ağaçlar, gradyan artırma kullanılarak inşa edilmiştir. Yani model, eğitim kaybını daha da aza indiren ağaçları kademeli olarak eklemektedir. İç düğümlere özellik testleri ekleyerek, kök düğümden başlayarak özyinelemeli olarak oluşturulur. Her iç düğümde, verilere bölünme uygulanarak elde edilen kazanca göre olası tüm özellik testleri değerlendirilir. En yüksek kazanım puanını geri veren test adayı daha sonra alınır ve her iki veri setinde de maksimum derinliğe ulaşılana veya kazanımlar belirli bir eşiğin altında kalıncaya kadar bölünür. Bir örnek tüm ağaçlardan geçirilerek ve ilgili yaprak skorları toplanarak bir tahmin hesaplaması yapılabilir. Karar ağacı yöntemlerinin topluluklarını kullanan gradyan artırma makinelerinin en önemli faydalarından biri, eğitimli bir tahmine dayalı modelden otomatik olarak özellik önemi tahminleri sağlayabilmeleridir. Gradyan artırmanın bir diğer faydası, güçlendirilmiş ağaçlar inşa edildikten sonra, her bir özellik için önem puanlarını almanın nispeten basit olmasıdır. Genel olarak önem, model içindeki güçlendirilmiş karar ağaçlarının yapımında her bir özelliğin

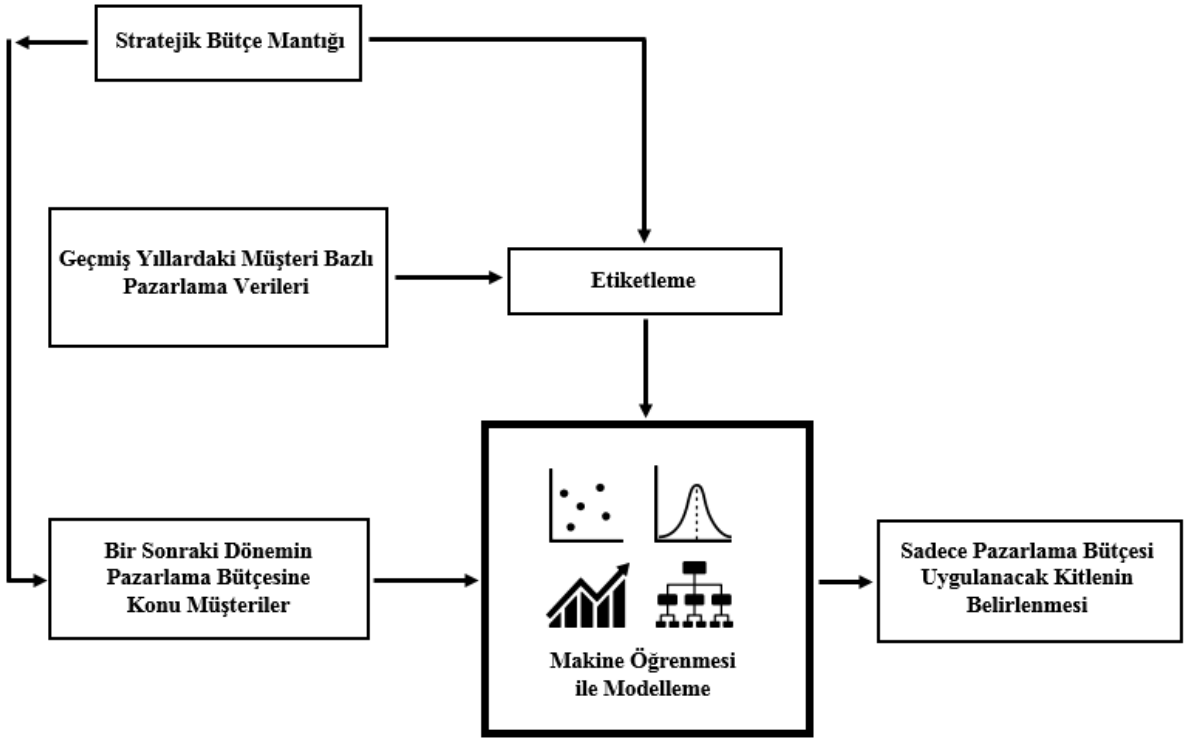
ne kadar yararlı veya değerli olduğunu gösteren bir puan sağlar. Karar ağaçlarıyla önemli kararlar almak için bir özellik ne kadar çok kullanılırsa, göreceli önemi o kadar yüksek olur. Bu önem, veri kümesindeki her bir özellik için açıkça hesaplanır ve özelliklerin sıralanmasına ve birbiriyle karşılaştırılmasına olanak tanır. Önem, tek bir karar ağacı için, her bir özellik bölme noktasının performans ölçüsünü iyileştirdiği miktara göre hesaplanır ve düğümün sorumlu olduğu gözlemlerin sayısına göre ağırlıklandırılır. Performans ölçüsü, bölünme noktalarını veya daha spesifik bir başka hata fonksiyonunu seçmek için kullanılan saflık (Gini indeksi) olabilir. Özellik önemlerinden sonra, model içindeki tüm karar ağaçlarının ortalaması alınır.

XGBoost algoritmasının pazarlama uygulamalarında kullanımına yönelik son yıllarda literatürde birkaç çalışmaya rastlanmıştır. Yoganarasimhan tarafından milyonlarca arama verisi üzerine yapılan çalışmada, her bir tüketici için kişiselleştirilmiş sıralaması, tüketicilerin arama sonuçlarına tıklama ve üzerinde durma olasılığını artırabileceğini göstermek için çoklu katkısal regresyon ağaçları (MART) kullanılmıştır (Yoganarasimhan, 2017). Kullanıcı geçmişi ve sorgu tipinin bir fonksiyonu olarak kişiselleştirmeden elde edilen getirilerdeki heterojenliği incelemek için tahmine dayalı bir model oluşturulmuştur. Diğer yandan, lojistik regresyonun taban çizgisi üzerinde hiçbir gelişme sağlanmadığı tespit edilmiştir. Rafieian ve Yoganarasimhan tarafından mobil uygulama reklamlarına yönelik bir hedefleme modeli oluşturulmasında XGBoost algoritmasından yararlanmışlardır (Rafieian ve Yoganarasimhan, 2020). Çalışmada, mobil uygulamalarda 27 milyondan fazla gösterimden elde edilen veri analiz edilmiştir. XGBoost algoritması, tüketicilerin mobil reklamlar için tıklama oranlarını tahmin etmede en küçük kareler (EKK) yöntemi, lojistik regresyon, LASSO regresyon ve rastgele ormanlar gibi yaygın olarak kullanılan diğer modellerden daha iyi performans gösterdiği tespiti yapılmıştır. Yine makalenin giriş bölümünde ayrıntıları verilen Jang (2019) ile Antipov ve Pokryshevskaya (2020)'nin yaptığı çalışmalar, büyük verilerin tahminine yönelik araştırmalarda GBM algoritmasının diğer algoritmalara göre daha ön plana çıktığını göstermektedir. Bu bağlamda, pazarlama alanında potansiyel müşteri grubunun tespitinin yapılmasına yönelik yapılan bu çalışmada tahmine dayalı bir model oluşturulması için kullanılan GBM algoritmasının çalışmanın amacına uygun bir makine öğrenmesi tekniği olduğunu desteklemektedir.

3.YÖNTEM

3.1. Veri Seti

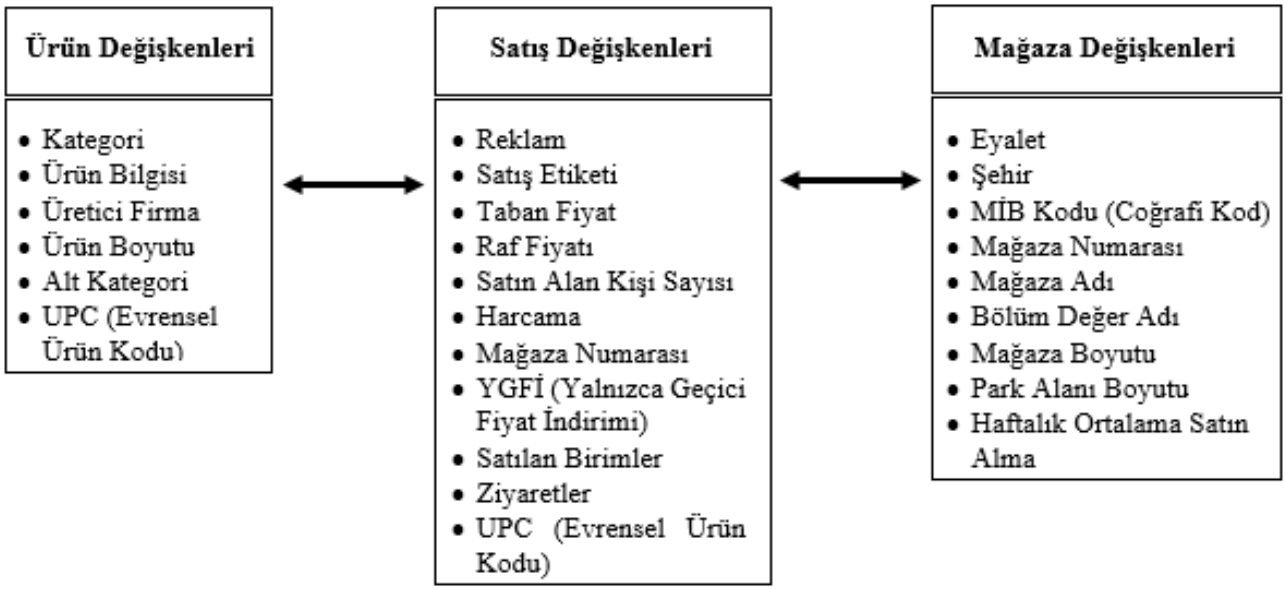
Pazarlamada kullanılacak olan bütçenin uygun müşteri grubuna harcanarak verimlilik sağlanmasının amaçlandığı bu çalışma kapsamında, Şekil 1 içerisinde gösterilen genel çerçeve uygulanmıştır. Bu çerçeve doğrultusunda, pazarlama için etkin olabilecek müşteri kitlesinin gradyan artırma algoritması aracılığıyla seçilmiş ve puanlanmıştır. Belirlenen çerçeve içerisinde Aralık 2011 ile Ocak 2019 tarihleri arasında toplanmış olan müşterilere ait pazarlama verileri üzerinden daha öncesinde pazarlamaya ayrılan ve başarı elde edilmiş olan potansiyel kitle işaretlenerek GBM algoritması ile modeller oluşturulmuştur. Model analizleri KNIME Analytics Platform 4.2.1 yazılımı ile yapılmıştır. Bu modellerin içinden en başarılısı seçilmiş ve bu modelde bir sonraki dönemde bütçeyi ilgilendirebilecek müşterilerin tamamı girdi olarak verilmiştir. Modelden geçen veriler etikenmiş ve puanlanmıştır. Bu sayede en yüksek puan ve başarı etiketini alan potansiyel müşterilerin belirlenmesi hedeflenmiştir.



Şekil 1. Çalışma için oluşturulan ve deney veri setine uygulanan genel çerçeve.

(General view of test data set for the study)

Markalara ve perakendecilere müşteri verileri üzerinden hizmet sağlayan Dunnhumby şirketinin bilimsel amaçlarda kullanılmak üzere sunmuş olduğu ve haftalık kahvaltı ürünlerinden elde edilen "Breakfast at the FRAT" başlığı altında toplanılan satış bilgileri çalışmanın deney veri setini oluşturmuştur (Dunnhumby, 2019). Dunnhumby şirketinin topladığı veriler 2011 ile 2019 yılları arasında 156 haftalık bir süreyi kapsamaktadır. Deney veri setinde seçilmiş olan 4 kategoriden en fazla satış rakamına ulaşan 3 markadaki en iyi 5 ürünün satış ve promosyon bilgileri yer almaktadır. Seçilmiş kategoriler donmuş pizza, kraker, ağız bakım suyu ve kahvaltılık gevrek olarak belirlenmiştir. Oluşturulan ham veri setindeki değişkenler Şekil 2 içerisinde sunulmuştur. Ürün, satış ve mağaza başlıkları hâlinde toplanılan ham veri dosyaları tek bir veri seti içerisinde birleştirilmiştir.



Şekil 2. Deney veri setindeki ürün, satış ve mağaza değişkenleri.

(Product,sales and store variables in test data set)

Dunnhumby şirketinin topladığı verilerde SKU ile ürün kategorisi sayısı kısıtlı olmasına karşın, verilerin gerçek dünyayla ilişkili olması, veri kalitesinin yüksek tutulması, dışa açık veriler içermesi, çok sayıda verinin bir araya getirilerek tahminde bulunulması nedeniyle bilimsel ve akademik açıdan oldukça kullanışlıdır ve veri bilimi için değerlidir. Bu veri setini değerli hâle getiren nedenlerden bir diğeri ürünlerin müşterilere gösterilip gösterilmediği veya tanıtılıp tanıtılmadığı ile ilgili reklam (display) değişkeniyle ilgili verilerin bulunmasıdır. Çünkü hedef olarak bu değişken seçilmiş ve oluşturulan model üzerinden verilerdeki örüntülere ulaşılmıştır. Bu değişken sayesinde yeni verilerin tahmini yapılarak puanlamalar yapılmıştır. Yine bu değişkene göre, deney setindeki 538643 veri içerisinden 479189 veri 0 şeklinde ve 59454 veri 1 şeklinde etiketlenmiştir. 0 ve 1 etiketleri, reklam görmeyenler ile reklam görenleri ifade etmektedir.

3.2. Özelliklerin Seçimi

Deney veri setindeki verilerin özellik seçimlerinde, bir özelliğin alt kümesini kullanarak modeli eğitmeye çalışan sarmalayıcı (wrapper) yöntemler kullanılmıştır. Sarmalayıcı yöntemler, daha önceki modellerden çıkarımlar yaparak, bu çıkarımlara bağlı olarak özellik alt kümesinden ekleme ya da kaldırma işlemi yapar (Xue vd., 2018). Çalışma içerisinde kullanılan sarmalayıcı yöntemlerden biri olan ileri özellik seçimi tekniği, başlangıçta özelliğin olmadığı tekrarlamalı bir yöntem olup her tekrarlamada yeni bir özelliğin eklendiği ve model performansında artış olmadığı müddetçe modeli geliştirecek olan özellikleri eklemeye devam eden bir yöntemdir. İleri özellik seçimi ile deney veri setinde yer alan 24 farklı özelliğe eklemeler yapılmış, verilerdeki karmaşıklık durumları en aza indirgenmiş ve en çok 0.91 doğruluk değerinin elde edilebildiği 19 farklı özelliğin seçimine imkân tanımıştır. Bu değişkenler; mağaza numarası, UPC, harcama, raf fiyatı, taban fiyat, satış etiketi, reklam, YGFİ, mağaza adı, eyalet, MİB kodu, bölüm değer adı, park alanı boyutu, mağaza boyutu, haftalık ortalama satın alma, ürün bilgisi, üretici firma, kategori, ürün boyutu şeklinde seçilmiştir. Çalışmada kullanılan diğer bir sarmalayıcı yöntem olan geri özellik eleme tekniğiyle modele bütün özellikler ile başlanılır ve her bir tekrarlamada en önemsiz olan özellikler elenir. Geri özellik eleme ile 8 farklı özelliğin (ziyaretler, raf fiyatı, satış etiketi, reklam, YGFİ, üretici firma, kategori, alt kategori) yer aldığı bir modelde en fazla 0.909 doğruluk değerine ulaşılmıştır.

3.3. Veri Analizi ve Sistem Mimarisinin Akışı

Deney setini kullanmadan önce, 3 farklı veri dosyası (satış, ürün, mağaza) birleştirilerek, ana dosya içerisindeki nümerik olarak boş olan kısımlar sıfır ile, kategorik özellikler ise en sık şekilde geçen özellikler ile doldurulmuştur. Verilerin analizinde karşılaşılan sorunlardan biri eksik verilerdir. Bu veriler, veri setinden çıkarılabilir veya yerleri farklı teknikler ile doldurulur. Eksik veriler ile çalışılmadan veri setinden doğrudan çıkarmak, yapılan analizin güvenirliğini azaltır. Eğer eksik verilerin değerleri rastgele şekilde oluşursa, o zaman bu değerler silinir. Bu nedenle, veri setinde ortaya çıkan eksikliğin veri yapısından kaynaklanan bir değer olup olmadığının bilinmesi gereklidir. Çalışma kapsamında ele alınan veri setinde yapısal bir eksikliğe rastlanmadığı için, eksik verilerin yerleri uygun değerler ile doldurulmuştur. Daha sonra, sistem performansını ölçmek için gelişigüzel şekilde 50 bin kayıt seçilmiş ve iki sarmalama yöntemi ile sistem denenmiştir. İleri özellik seçiminde elde edilen sonucun doğruluk değeri daha iyi olduğundan 24 farklı özelliğin yer aldığı model eğitim ve test veri seti olarak ikiye ayrılmıştır. Bu işlem, modelin doğruluk ve güvenirliğini belirlemek için kullanılır. Makine öğrenmesi algoritmalarının performans değerlendirmesi için kullanılacak olan veri setlerinin fonksiyonları önemlidir. Eğitim veri seti modelin eğitimi için kullanılırken, test veri seti genelleme hatalarını tahmin etmede kullanılır. Test veri seti için etiketli verilere gereksinim duyulmaktadır. Çalışma kapsamında, 538643 verinin %70'lik kısmı modelin eğitiminde ve %30'luk kısmı modelin test edilmesinde kullanılmıştır.

4.BULGULAR

Çalışma içerisinde oluşturulan modelin başarısının test edilmesinde kullanılan GBM algoritması ile analizi KNIME 4.2.1 yazılımı aracılığıyla gerçekleştirilmiştir. Model performansının ölçümü için birçok sınıflandırma ölçütünü hesaplayabilen karmaşıklık matrisi yöntemi kullanılmıştır. Karmaşıklık matrislerinde sınıflandırılan veriler gerçek ve tahmin sınıfı olarak ikiye ayrılır. Sınıflandırmanın sonuçları matrisin satırlarında yer alırken gerçek değerleri matrisin sütunlarında yer alır. Deney veri seti içerisinde hedef değişken veya özellik olarak

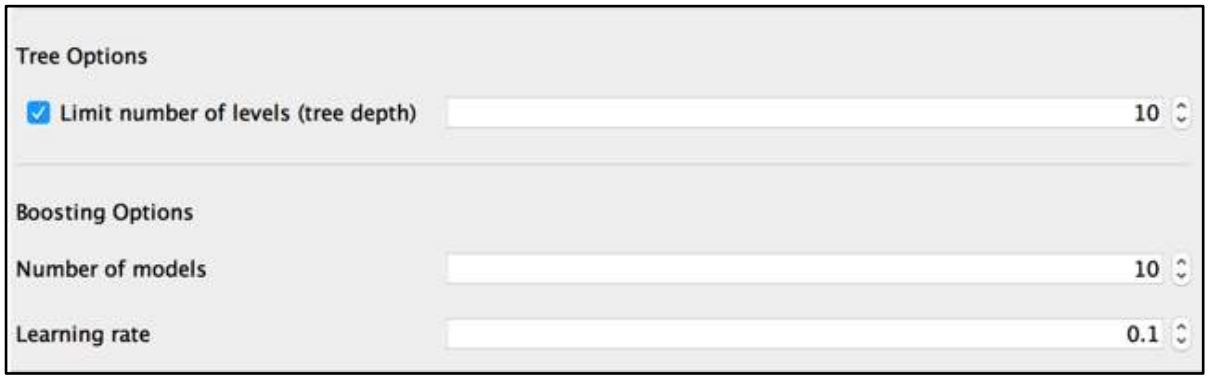
seçilen reklam değişkeni için satılan ürünlerin müşterilere gösterilip gösterilmemesi durumları önem arz etmektedir. Tablo 1 içerisinde ikili sınıflandırma durumunda reklam gören müşteriler ile reklam görmeyen müşterilere ait karmaşıklık matrisi açık bir şekilde gösterilmiştir. Ayrıca gradyan artırma algoritması için modelin doğruluğu, duyarlılığı, kesinliği, seçiciliği, F-Ölçütü ve AUC değeri gibi ölçütleri (metrik) hesaplanmıştır. Bu ölçütlerin her biri 0 ile 1 arasında değerler almaktadır. Ölçüt değeri 1 değerine ne kadar yakında sınıflandırma bakımından yüksek bir performans gösterdiği anlamına gelir.

Tablo 1. Reklam değişkenine ait karmaşıklık matrisi gösterimi.

(Complexity matrix view belonging to display variable)

Karmaşıklık Matrisi		**Verinin Tahmin Edilen Sınıfı**	
		Reklam Gören	**Reklam Görmeyen**
Verinin Gerçek Sınıfı	Reklam Gören	Doğru Pozitif (DP)	Yanlış Negatif (YN)
	Reklam Görmeyen	Yanlış Pozitif (YP)	Doğru Negatif (DN)

GBM algoritması modele uygulanılırken, model sayısı 10 olarak belirlenmiştir, bu sayı ağacın 10 defa tekrarlanması anlamını taşımaktadır. Tahmin hatalarını düzeltmek için ağaç derinliği sayısı 10 ile sınırlı tutulmuştur. Hatanın her ağaçtan diğerine ne kadar hızlı düzeltildiğine karşılık gelen öğrenme oranı 0.1 olarak belirlenmiştir. GBM algoritmasındaki her yinelemeli adımda eklenen karar ağacının katkısı, öğrenme oranı adı verilen bir küçülme parametresi kullanılarak hesaplandığında daha iyi sonuçlar verir. Çünkü GBM algoritmasında öğrenmenin hızlı olması ve eğitim verilerine fazla uyması karar ağaçlarıyla ilgili bir sorunu ortaya çıkarır. GBM açısından bir küçülme prosedürü uygulanmasındaki ana fikir, daha fazla sayıda küçük adımın öğrenmeyi yavaşlatması ile daha az sayıda büyük adımdan daha yüksek bir doğruluk sağlamasıdır (Friedman, 2001, s. 1208; Touzani vd., 2017, s. 1537). Öğrenme oranı 0 ile 1 arasında ($0 < LR \leq 1$) bir değer alabilir ve ne kadar küçük olursa model o kadar doğru olur. Bununla birlikte, daha güçlü bir küçülme oranının seçilmesi sonucu öğrenme değerinin yineleme sayısıyla ters orantılı olacağından, yakınsama elde etmek için daha fazla yineleme olacağı anlamına gelir. Bu nedenle, çalışmada daha fazla yinelemenin tercih edildiği küçük bir öğrenme oranı tercih edilmiştir. Şekil 2 içerisinde veri setine uygulanan GBM algoritmasına ait giriş parametrelerine yer verilmiştir.



Şekil 2. Test veri setine uygulanan GBM algoritmasına ait giriş parametreleri.

(GBM algorithm parameters executed to test data set)

GBM algoritmasının modele uygulanması neticesinde, reklam değişkeninin performasını belirlemek için kullanılan ikili sınıflandırma şeklindeki karmaşıklık matrisi değerleri eğitim veri seti için Tablo 2'de ve test veri seti için Tablo 3'te sunulmuştur. Pozitif ve negatif verileri gösteren ve test veri setinin sınıflandırma ölçüsü

olan karmaşıklık matrisinin sonuçları GBM algoritmasının değerlendirilebilmesi açısından önem taşımaktadır.

Tablo 2. GBM algoritmasının ikili karmaşıklık matrisine ilişkin eğitim verisi için performans değerleri.

(Performance variables for training data of GBM algorithm double comlexity matrix)

Algoritma Türü	Karmaşıklık Matrisi		Verinin Tahmin Edilen Sınıfı		Toplam Veri Sayısı
			Reklam Gören	Reklam Görmeyen	
Gradyan Artırma Makineleri	Verinin Gerçek Sınıfı	Reklam Gören (Display = 1)	27401	14217	41618
		Reklam Görmeyen (Display = 0)	5731	329701	335432
	Toplam Veri Sayısı		33132	343918	377050

Tablo 2 içerisinde ayrıntıları ile verilen karmaşıklık matrisi yöntemine ilişkin performans değerleri göz önüne alındığında; toplam 377050 verinin işleme alındığı, DP, YP, YN ve DN değerlerine ait veri sayısının sırasıyla 27401, 5731, 14217 ve 329701 olduğu görülmüştür. 357102 veri, doğru şekilde sınıflandırılan veri sayısıdır. 14217 veri gerçekte reklam görüp sınıflandırmanın sonucunda reklam görmeyen olarak, gerçekte reklam görmeyip sınıflandırmanın sonucunda reklam gören olarak etikenmiştir.

Tablo 3. GBM algoritmasının ikili karmaşıklık matrisine ilişkin test verisi için performans değerleri.

(Performance variables for test data of GBM algorithm double comlexity matrix)

Algoritma Türü	Karmaşıklık Matrisi		Verinin Tahmin Edilen Sınıfı		Toplam Veri Sayısı
			Reklam Gören	Reklam Görmeyen	
Gradyan Artırma Makineleri	Verinin Gerçek Sınıfı	Reklam Gören (Display = 1)	11322	6514	17836
		Reklam Görmeyen (Display = 0)	2743	141014	143757
	Toplam Veri Sayısı		14065	147528	161593

Tablo 3 içerisinde sunulan karmaşıklık matrisi yöntemine ilişkin test verilerine ait performans değerleri incelendiğinde; toplam 161593 verinin işleme alındığı, DP, YP, YN ve DN değerlerine ait veri sayısının sırasıyla 11322, 2743, 6514 ve 141014 olduğu görülmüştür. Doğru şekilde sınıflandırılan veri sayısı 152336'dır. 6514 verinin ise gerçekte reklam görüp sınıflandırmanın sonucunda reklam görmeyen olarak, gerçekte reklam görmeyip sınıflandırmanın sonucunda reklam gören olarak etikenmiştir.

Eğitim ve test verilerinin model performansının ölçümüne ait metriklere sırasıyla Tablo 4 ve Tablo 5 içerisinde yer verilmiştir.

Tablo 4. GBM algoritması için eğitim veri setinin model performansı ölçümüne ait ölçüt hesaplamaları.

(Model performance measurement results of training data set for the GBM algorithm)

Algoritma Türü	Model Performansının Ölçümüne Ait Hesaplamalar					
		Duyarlılık	Keskinlik	Seçicilik	F-Ölçütü	Doğruluk
Gradyan Artırma Makineleri	Reklam Gören	0.658	0.827	0.983	0.733	
	Reklam Görmeyen	0.983	0.959	0.658	0.970	
	Genel	0.821	0.893	0.821	0.852	0,947

Tablo 4'te verilen GBM algoritması için eğitim veri setinin model performansı ölçümüne ait ölçüt hesaplamalarından doğruluk değerinin 0.947 olduğu görülmektedir. Eğitim veri setinden doğru olarak tahmin edilen verilerin oranının yüksek bir performans değerinde olduğu görülmektedir. Doğruluk, bir bütün olarak modelin doğru şekilde sınıflandırılma oranıdır. Tablo 4'e göre; sınıflandırıcının gerçekte reklam görenleri hangi oranda reklam gören şeklinde tahminde bulunduğuyla ilgili duyarlılık değeri 0.821, gerçekte reklam görenlerin sayısının sınıflandırıcının reklam gören şeklinde tahminde bulunduğu sayıya bölünmesi ile elde edilen kesinlik değeri 0.893, sınıflandırıcının gerçekte reklam görmeyenleri hangi oranda reklam görmeyen şeklinde tahminde bulunduğuyla ilgili seçicilik değeri 0.821, yanlış negatif ve pozitif değerlerin hesaplandığı, kesinlik ile duyarlılık değerlerinin harmonik ortalamasını veren ve bu iki ölçütün yalnız başına eksik olabileceği göz önüne alındığı için hesaplanan F-Ölçütü değeri 0.852'dir. GBM algoritması için eğitim veri setinin model performansı ölçümüne ait ölçütlerin yalnızca reklam görenler açısından incelendiğinde, duyarlılığın 0.658, kesinliğin 0.827, seçiciliğin 0.983 ve F-Ölçütünün 0.733 değerlerini verdiğini; reklam görmeyenler açısından ise duyarlılığın 0.983, kesinliğin 0.959, seçiciliğin 0.658 ve F-Ölçütünün 0.970 değerlerini vermiştir. Reklam görmeyen açısından duyarlılığın, kesinliğin ve F-ölçütünün, reklam gören açısından yalnızca seçiciliğin 1 değerine en yakın oranda yüksek bir model performansı sergilediği görülmüştür.

Tablo 5. GBM algoritması için test veri setinin model performansı ölçümüne ait ölçüt hesaplamaları.

(Model performance measurement results of test data set for the GBM algorithm)

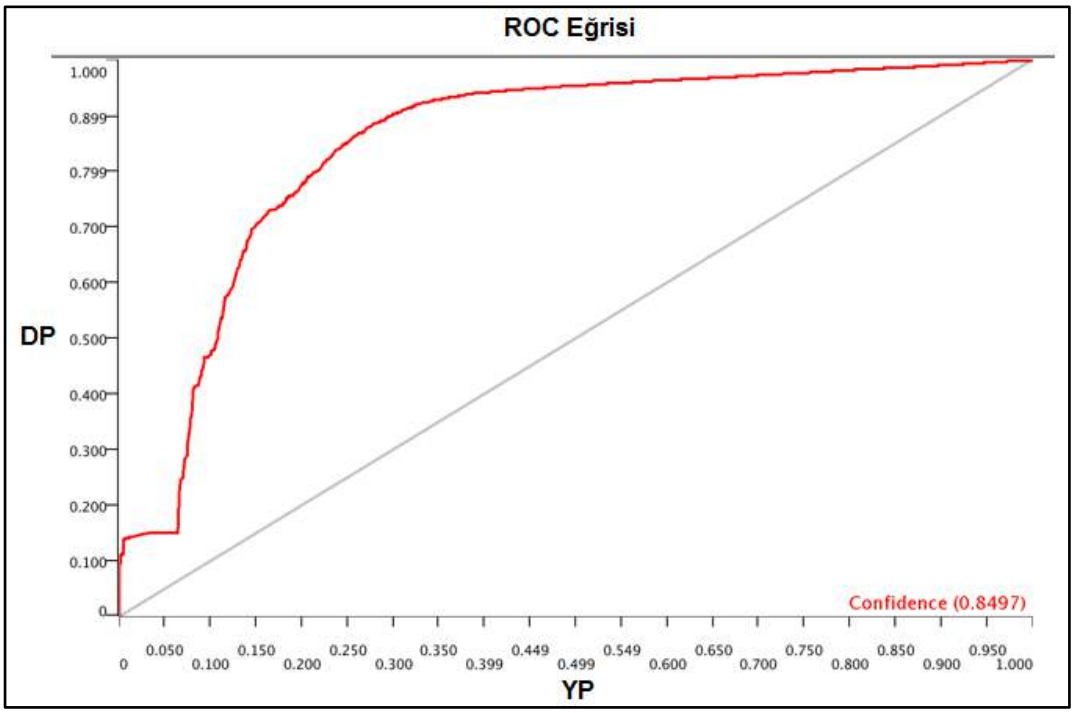
Algoritma Türü	Model Performansının Ölçümüne Ait Hesaplamalar					
		Duyarlılık	Keskinlik	Seçicilik	F-Ölçütü	Doğruluk
Gradyan Artırma Makineleri	Reklam Gören	0.635	0.805	0.981	0.710	
	Reklam Görmeyen	0.981	0.956	0.635	0.968	
	Genel	0.808	0.880	0.808	0.839	0.943

Tablo 5 içerisinde ayrıntıları ile verilen GBM algoritması için test verisinin model performansı ölçümüne ait ölçüt hesaplamalarına ilişkin değerler göz önüne alındığında; doğruluk değerinin 0.943 olduğu görülmektedir. Tablo 5'te genel hesaplamalara doğrultusunda; duyarlılık değeri 0.808, kesinlik değeri 0.880, seçicilik değeri 0.808, F-Ölçütü değeri 0.808'dir. Model performansı ölçümüne ait ölçütlerin yalnızca reklam görenler açısından değerlendirildiğinde, duyarlılığın 0.635, kesinliğin 0.805, seçiciliğin 0.981 ve F-Ölçütünün 0.710 değerlerini verdiğini; reklam görmeyenler açısından ise duyarlılığın 0.981, kesinliğin 0.956, seçiciliğin 0.635 ve F-Ölçütünün 0.710 değerlerini vermiştir. Eğitim veri setinde olduğu gibi test veri setinde de reklam görmeyenler açısından duyarlılığın, kesinliğin ve F-ölçütünün, reklam görenler açısından yalnızca seçiciliğin 1 değerine en yakın oranda yüksek bir model performansı sergilemiştir.

Tablo 4 ve Tablo 5'te sunulan eğitim ve test veri setlerine ait değerlerin birbirlerine çok yakın ve uyumlu olduğu görülmektedir. Sınıflandırma modelinin oluşturulmasında yararlanılan test tekniğinin önemini bu çalışmada öne çıkarmaktadır. Eğitim ve test veri setleri için sınıflandırıcı performansından elde edilen doğruluk değerinin birbirine çok yakın değerlerde olması, doğruluk değerinin gerçek sonucu yansıttığını ve bu değerin kullanılması gerektiği sonucunu ortaya çıkarmıştır. Ayrıca, GBM algoritmasının modele uygulanması sonucu elde edilen performans ölçüt değerlerinin yüksek başarı sağlaması, kullanılan algoritmanın model için uygunluğunu ve deney veri setindeki satış verilerinin GBM algoritması kullanılarak analiz edilmesinin doğruluğunu ortaya koymaktadır. Bu sonuçlar, yapılan çalışmanın son yıllarda literatürde tahmine dayalı satış ve promosyon verileri üzerine yapılan diğer çalışmalarda (Jang, 2019; Antipov ve Pokryshevskaya, 2020) kullanılan GBM algoritmasının geleneksel algoritmalara karşı üstünlüğünü de ortaya çıkarmıştır.

Çalışma kapsamında yararlanılan GBM algoritmasının sınıflandırma tahminin belirlenmesi amacıyla sınıflandırma analizi için tahminin ne oranda sıralanmasının bir ölçütü olan AUC (Area Under Curve) kriteri de göz önünde bulundurulmuştur. Bu bağlamda, AUC kriteri örneklerinden biri olan ROC eğrisi algoritmanın model performans değerlendirmesi için hesaplanmıştır. AUC krtiteri de 0 ile 1 arasında değerler almaktadır

ve sınıflandırma için en iyi performansı sergileyen değer 1'e en yakın değerlerdir. Son yıllarda makine öğrenmesi uygulamalarından önem kazanan ROC (Receiver Operating Characteristics) eğrileri; tanısal kararlar verilmesinde, maliyet analizlerinde ve birçok alanda model performansının durumunu değerlendirmek için sıklıkla kullanılmaktadır. Çalışma kapsamında yararlanılan GBM algoritmasına yönelik eğitim ve test veri seti için AUC kriterine ait değerler sırasıyla 0.861 ve 0.850 olarak hesaplanmıştır.



Şekil 3. GBM algoritması için test veri setine ait ROC eğrisi.

(ROC curve belongs to test data for GBM algorithms)

5. SONUÇ VE TARTIŞMA

Bu çalışma, makine öğrenmesi algoritmalarının işletmelerdeki kaynak kullanımına direkt katkısı hakkında somut bir örnek oluşturmaktadır. Bir işletmenin pazarlama bütçesini tüm müşterilerine harcamak yerine sadece potansiyel müşteri kitlesine harcanmasına imkân tanıyan bir algoritma kullanılarak pazarlamanın daha etkin ve verimli olabileceği müşterilerin belirlenmesine yönelik çalışma kapsamında oluşturulan model, seçilen özellikler üzerinden müşteri davranışlarının daha iyi tahmin edilmesi ve pazarlama bütçesiyle ilgili yönetsel kararlara yardımcı olması için ortaya konulmuştur. Bir işletmenin doğru bir satış tahmini ve öngörüye sahip olmaması, alınan birçok iş kararında güven vermeyen tahminlere göre plan ve stratejilerin belirlenmesine yol açacak ve bu durum işletme bütçesinde maliyet artışına, kâr oranlarının azalmasına ve sektörde çok sayıdaki fırsatın elden kaçmasına neden olacaktır. Dikkatle oluşturulmuş bir satış tahmini ve öngörüsü, işletmelerin gelecekleri adına daha tutarlı kararlar almasında önemli bir rol oynayacaktır. Satış verileri ile ilgili doğru ve tutarlı tahminlerin yapılmasında ve bu verilerin analizleri sonucu zaman, maliyet ve kâr analizlerinin yapılmasında makine öğrenmesi algoritmalarının katkısı büyüktür

Yapay sinir ağları ve modelleme çalışmalarında da (Aker vd 2022) finansal başarısızlığın tahmin edilmesinin finans alanındaki önemli araştırma konularından birisi olduğu,finansal başarısızlık tahmininin amacının, çeşitli ekonometrik göstergelerden yararlanarak bir işletmenin finansal durumunu tahmin etmeyi sağlayacak bir tahmin modeli geliştirmek olduğu anlatılmıştır. İşletmenin alacaklıları ve yatırımcılar açısından bir işletmenin finansal açıdan başarısız olma ihtimalini bilmek, alınacak kararlar öncesi ürün snıflandırması ve fiyatlama gibi konularda oldukça önemli bir konu haline geldiği anlatılmıştır.

Model analizleri ve uygulama açısından özelliklerin önem analizi sonrası özellik sayısını az sayıda tutarak yüksek bir doğruluğa ulaşılması, veriler içerisinde en uygun ve önem seviyesi yüksek olan özelliklerden seçilerek oluşturulan modelin veri bilimi açısından değerli olduğu anlamını taşımaktadır. Ulusal ve Uluslararası, bu ve benzeri verimlilik ile ilgili çok kısıtlı sayıda çalışmaya ulaşılabildiğinden, yorumlar sınırlı kalmaktadır. Daha genellenebilir yorumlar yapabilmemize olanak sağlayacak şeffaf çalışmalara ihtiyaç duyulmaktadır ancak firmaların bu tür bilgileri paylaşma oranları gizlilik ve teknolojik hırsızlık gibi endişelerle, giderek azalmaktadır. Araştırmacılara bu konu hakkında neden farklı örneklem gruplarıyla çalışma gerçekleştiremedikleri, daha fazla bilgiye ulaşacak kaynaklarla ilgili çalışmaları neden yapamadıkları sorulmalı, bilgi sağlayacak firmalarla güvenlik endişelerini giderecek altyapılara dayalı, paylaşımcı sistemler üzerinde tartışılmalıdır.
Bu tür analizlerin sayısının artması hem yeni dönem planlamaları için çalışan tüm sektörlerdeki firmalara, hem de gerçek veriler üzerinden farklı algoritmaları geliştirebilme imkanını yakalayacak araştırmacılara fayda sağlayacaktır. Doğru kanallar üzerinden veri paylaşılmasının nasıl destekleneceği, hangi yöntemlerle güvenlik endişelerin asgari orana ineceği bugünün ve yarının en önemli tartışma konusudur.

Çalışma kapsamında birkaç kısıtlılık söz konusudur. Bu kısıtlılıkların ilki, sadece bir perakende şirketinin topladığı satış verilerinin analiz edilebilmiş olmasıdır. Dünya çapında akademik açıdan güvenli ve kullanışlı veri toplayabilen ve bu verileri halka açık kaynak olarak sunabilen şirket sayısı oldukça sınırlıdır. Çalışmanın ikinci bir kısıtlılığı, Dunnhumby şirketi tarafından elde edilen satış verilerinin belirli dönemlerde, belirli özelliklerde ve belirli markalarda toplanmış olmasıdır. Veri sayısının artırılması, yeni özelliklerin eklenmesi ve marka sayısının daha fazla sayıda tutulması durumunun GBM algoritmasının performansına nasıl bir etki edeceği konusunda bir belirsizlik oluşturmaktadır. Çalışmada ürünlerin müşterilere gösterilip gösterilmeme durumu ile ilgili reklam özelliğinin esas alınarak her bir özelliğin seçiminde ileri özellik seçimi yöntemi uygulanmış ve bu sayede karmaşıklık durumları en aza indirgenerek 24 farklı özelliğin kullanıldığı deney veri setinde yüksek bir doğruluk değeri elde edilmiştir. Gerçek dünya sistemine uygulanarak ortaya konulan bu modelin performans ölçümünden elde edilen değerlerin yüksek sonuçlar verdiği ve ilgili alanda bu modelin kullanılması tavsiye edilmektedir.

GBM algoritmasının performasına ilişkin doğruluk ölçütüne ait %94'lük bir başarı değeri, modelin sınıflandırma performasının çok iyi bir düzeyde olduğunu ortaya koymaktadır. GBM algoritmasının müşterilerin satış tepkisinde iyi sonuçlar vermesi, bu algoritmanın çalışmanın amacına uygun olduğunu göstermektedir. Bu bağlamda, pazarlamanın daha etkin ve verimli olabileceği müşterilerin belirlenmesinde ve bütçe verimliliği sağlanması için kullanılması önerilmektedir. Bu model hata teriminin artırımlı olarak en aza indirilmesi yoluyla tahmin gücünün doğruluğunu artıran kolektif bir algoritmadır. İlk temel öğrenen (çoğunlukla bir ağaç) yetiştirildikten sonra, serideki her ağaç, hatayı azaltmak amacıyla önceki ağaçlardan gelen tahminin «sözde artıklarına" uygun hareket eder (Brown ve Mues, 2012). Yine 2022 yılında yapılan başka bir çalışmada GBM algoritmasının kredi skorlanmasında da yüksek oranda başarılı olduğu ortaya koyulmuştıur (Altan, G. ve Demirci, S. 2022).

Satışa ve promosyona dayalı büyük verilerin analiz edilerek pazarlama için etkin olabilecek müşteri kitlesinin gradyan artırma algoritması aracılığıyla belirlenmesine yönelik gerçekleştirilen bu çalışmada uygulanan adımlar ve ortaya konulan model bütçe verimliliği açısından örnek bir çalışma niteliği taşımaktadır. Gerçek veriler kullanılarak yapılan bu çalışma, önerilen modelin ilgili alana sağladığı avantajı ortaya çıkarmıştır. Model, bir yineleme sırasında GBM algoritmasının paralel ve bağımsız kullanımına dayanmaktadır. Aslında, GBM algoritması tek bir özellik ile ilgilenir. Bununla birlikte, önerilen algoritmanın mimarisi, özelliklerin kombinasyonlarını dikkate alabilecek bir model geliştirme fikrine yol açar. Bu durumda, özelliklerin korelasyonu dikkate alınarak çeşitli çalışmalar yapılabilir. Ayrıca bütçe verimliliği üzerinde makine öğrenmesi teknikleri bu çalışmada gösterildiği şekilde uygulanabileceği gibi, gelecek çalışmalar adına müşteri verilerinin daha verimli kullanılmasının işletmelere zaman, maliyet ve kâr açısından nasıl bir fayda sağlayabileceği ile ilgili uygulamalı bir çalışma yapılabilir. Bu çalışmada oluşturulan modelin bir işletmeye uygulanmadan önce ve uyguladıktan sonraki durumlarının karşılaştırılmasına yönelik bir araştırma yapılabilir.Zira büyük boyutlu ham veriler üzerinde onlarca kişinin çok uzun zaman ve emek harcadığı bir analizi, birkaç kişinin denetimdeki makine öğrenmesi teknikleri ile otomatik olarak daha kısa sürede ve daha yüksek bir doğrulukla elde edebilmek mümkündür.

Makine öğrenmesi teknikleri ile bir işletmenin bütçeleme ve finansal planlama ile ilgili yatırımlarının en üst seviyeye taşınması hedeflenir. Bu nedenle, makine öğrenmesi tekniklerinin işletmelere hem maliyet hem de zaman açısından büyük avantajlar sağladığı açıktır. Diğer taraftan, makine öğrenmesi teknikleri varsayımları ortadan kaldırarak işletmelere optimize çözüm önerileri sunabilmektedir. Bu bağlamda, doğru iş kararlarının alınmasında, yapılan işlerde süreklilik sağlanmasında ve finansal kaynaklardan verimlilik elde edilmesinde makine öğrenmesi tekniklerinin tüm işletmelerde yaygınlaşması ve kullanılması gerekmektedir.

Sonuç olarak, yapılan çalışma aracılığıyla bilgisayar ve veri bilimleri aracılığıyla pazarlama alanına katkıda bulunulması ve öncü bir makine öğrenmesi modeli ortaya konulması amaçlanmıştır. Çalışma kapsamında kurgulanan modelin işletmelerin yalnızca potansiyel müşterilerine pazarlama maliyet ayırarak bütçe verimliliği sağlaması açısından önemlidir. Gelecekte yapılması planlanan çalışmalarda hedef kitle analizi ve davranış alışkanlıklarında yapay zeka ve akine öğrenmesi en sık rastlanan metod olacaktır. Çalışma kapsamının ve altyapısının geliştirilerek ortaya konulacak olan modellerin işletmelerin bütçe verimliliğine katkı sağlaması en büyük temennidir.

KAYNAKÇA

Ailawadi, K. L., Harlam, B. A., César, J. ve Trounce, D. (2007). Practice prize report: quantifying and ımproving promotion effectiveness at CVS. *Marketing Science*, 26, 566-575. https://doi.org/10.1287/mksc.1060.0245

Ali, Ö. G., Sayın, S., Van Woensel, T. ve Fransoo, J. (2009). SKU demand forecasting in the presence of promotions. *Expert Systems with Applications*, 36, 12340-12348. https://doi.org/10.1016/j.eswa.2009.04.052

Aker, Yusuf; Karavardar, Alper. Mali Cözüm Dergisi; Istanbul Vol. 32, (Jan/Feb 2022): 151-169.

Altan, G. Ve Demirci, S. (2022). Makine öğrenmesi ile nakit akış tablosu üzerinden kredi skorlaması: XGBoost yaklaşımı.İktisat PolitikasıAraştırmaları Dergisi, 13(3), 397-424. https://doi.org/10.26650/JEPR1114842

Antipov, E. A. ve Pokryshevskaya, E. B. (2020). Interpretable machine learning for demand modeling with high-dimensional data using Gradient Boosting Machines and Shapley values. *Journal of Revenue and Pricing Management*, 1-10. https://link.springer.com/article/10.1057/s41272-020-00236-4

Bradlow, E. T., Gangwar, M., Kopalle, P. ve Voleti, S. (2017). The role of big data and predictive analytics in retailing. *Journal of Retailing*, 93, 79-95. https://doi.org/10.1016/j.jretai.2016.12.004

Brown, L. & Mues, C. (2012). An experimental comparison of classification algorithms for imbalanced credit scoring data sets. Expert Systems with Applications, 39, 3446-3453

Bulut, F. (2019). Bankacılık Sektöründe Makine Öğrenmesi Yöntemleriyle Müşteri İlişkileri Yönetiminin Zenginleştirilmesi. *Avrupa Bilim ve Teknoloji Dergisi*, (16), 382-394. https://doi.org/10.31590/ejosat.520295

Chandukala, S. R., Edwards, Y. D. ve Allenby G. M. (2011). Identifying unmet demand. *Marketing Science*, 30(1), 61-73. https://doi.org/10.1287/mksc.1100.0589

Chen, T. ve Guestrin, C. (2016). Xgboost: A scalable tree boosting system. In: Proc. of the 22nd SIGKDD Int. Conf. on Knowledge Discovery and Data Mining. *ACM (2016)*, 785-794. https://dl.acm.org/doi/pdf/10.1145/2939672.2939785

Cui, D. ve Curry, D. (2005). Prediction in marketing using the support vector machine. *Marketing Science*, 24(4), 595-615. https://www.jstor.org/stable/40056988

Cui, G., Wong, M. L. ve Lui, H.-K. (2006). Machine learning for direct marketing response models: Bayesian networks with evolutionary programming. *Management Science*, 52(4), 597-612. https://doi.org/10.1287/mnsc.1060.0514

Dimitrieska S., Stankovska A. ve Efremova T. (2018). Artifical intelligence and marketing. *Entrepreneurship*, 6(2), 298-304. https://doi.org/10.1177/14707853211018428

Dunnhumby (2019). A Time Series Analysis: Breakfast at the Frat. https://www.dunnhumby.com/source-files/, Erişim Tarihi: 01.04.2021.

Dzyabura, D. ve H. Yoganarasimhan (2018). *Machine learning and marketing*. In: Handbook of Marketing Analytics, Edward Elgar Publishing. ISBN: 9781784716745

Ferreira, K. J., Lee, B. H. A. ve Simchi-Levi, D. (2015). Analytics for an online retailer: Demand forecasting and price optimization. *Manufacturing & Service Operations Management*, 18, 69-88. https://doi.org/10.1287/msom.2015.0561

Friedman J. H. (2001). Greedy function approximation: A gradient boosting machine. *Annals of Statistics*, 29, 1189-1232. https://www.jstor.org/stable/2699986

Gedenk, K. (2018). *Retailer promotions*. In Handbook of Research on Retailing, ed. K. Gedenk. Cheltenham: Edward Elgar Publishing. ISBN: 9781786430274

Grushka-Cockayne, Y., Jose, V. R. R. ve Lichtendahl K. C. (2017). Ensembles of overfit and overconfident forecasts. *Management Science*, 63(4), 1110-1130. https://doi.org/10.1287/mnsc.2015.2389

Gürsakal, N. (2017). Makine öğrenmesi ve derin öğrenme. Bursa: Dora Basım.

Jacobs, B. J. D., Donkers, B. ve Fok D. (2016). Model-based purchase predictions for large assortments. *Marketing Science*, 35(3), 389-404. https://doi.org/10.1287/mksc.2016.0985

Jang, H. (2019). A decision support framework for robust r&d budget allocation using machine learning and optimization. *Decision Support Systems*. 121, 1-12. https://doi.org/10.1016/j.dss.2019.03.010

Jordan, M. I. ve T. M. Mitchell (2015). Machine learning: Trends, perspectives, and prospects. *Science*, 349(6245), 255-260. https://www.science.org/doi/10.1126/science.aaa8415

Kaynar, O., Tuna, M. F., Görmez, Y. ve Deveci, M. A. (2017). Makine öğrenmesi yöntemleriyle müşteri kaybı analizi. *Cumhuriyet Üniversitesi İktisadi ve İdari Bilimler Dergisi*, 18(1), 1-14. http://esjournal.cumhuriyet.edu.tr/tr/pub/issue/32216/357723

Liu, Q. ve Wu, Y. (2012). *Supervised Learning*. In: Seel N.M. (eds) Encyclopedia of the Sciences of Learning. Springer, Boston, MA. https://doi.org/10.1007/978-1-4419-1428-6

Ma, S. ve Fildes, R. (2017). A retail store sku promotions optimization model for category multi-period profit maximization. *European Journal of Operational Research*, 260, 680-692. https://doi.org/10.1016/j.ejor.2016.12.032

Ma, S., Fildes, R. ve Huang, T. (2016). Demand forecasting with high dimensional data: The case of SKU retail sales forecasting with ıntra-and ınter-category promotional information. *European Journal of Operational Research*, 249, 245-257. https://doi.org/10.1016/j.ejor.2015.08.029

MacKenzie, I., Meyer, C. ve Noble, S. How retailers can keep up with consumers.4-6. McKinsey & Company. https://www.mckinsey.com/industries/retail/our-insights/how-retailers-can-keep-up-with-consumers, Erişim Tarihi: 21.12.2021.

Onan, A. (2017). Twitter mesajlari üzerinde makine öğrenmesi yöntemlerine dayalı duygu analizi. *Yönetim Bilişim Sistemleri Dergisi*, 3(2), 1-14. https://doi.org/10.17780/ksujes.819367

Rafieian, O. ve Yoganarasimhan, H. (2020). Targeting and Privacy in Mobile Advertising. *Marketing Science*, 40(2), 193-218. https://doi.org/10.1287/mksc.2020.1235

Singh H. (2018). Understanding gradient boosting machines. https://towardsdatascience.com/understanding-gradient-boosting-machines-9be756fe76ab, Erişim Tarihi: 20.12.2021.

Sun, Z.-L., Choi, T.-M., Au, K.-F. ve Yu., Y. (2008). Sales forecasting using extreme learning machine with applications in fashion retailing. decision support systems, 46, 411-419. https://doi.org/10.1016/j.dss.2008.07.009

Tam, K. Y. ve M. Y. Kiang (1992). Managerial applications of neural networks: The case of bank failure predictions. *Management Science*, 38(7), 926-947. http://dx.doi.org/10.1287/mnsc.38.7.926

Touzani, S., Granderson, J. ve Fernandes S. (2018). Gradient boosting machine for modeling the energy consumption of commercial buildings. Energy and Buildings, 158(6), 1533-1543. https://doi.org/10.1016/j.enbuild.2017.11.039

Tuna, M. F., Kaynar, O. Ve Akdoğan, M. Ş. (2021). Otellere ilişkin çevrimiçi geribildirimlerin makine öğrenmesi yöntemleriyle duygu analizi. *İşletme Araştırmaları Dergisi*, 13(3), 2232-2241. https://doi.org/10.20491/isarder.2021.1258

Xue, X., Yao, M. ve Wu, Z. (2018). A novel ensemble-based wrapper method for feature selection using extreme learning machine and genetic algorithm. *Knowledge and Information Systems*, 57, 389-412. https://link.springer.com/article/10.1007/s10115-017-1131-4

Yang, D. ve Zhang, A. N. (2018). *Forecast UPC-level FMCG demand*, Part IV: Statistical Ensemble. In: 2018 IEEE International Conference on Big Data (Big Data), 3180-3185. https://ieeexplore.ieee.org/document/8622029

Yoganarasimhan, H. (2019). Search Personalization using Machine Learning. *Management Science*, 66(3), 1045-1070. https://doi.org/10.1287/mnsc.2018.3255

Yu, A. (2019). How Netflix Uses AI, Data Science, and Machine Learning-From A Product Perspective. https://becominghuman.ai/how-netflix-uses-ai-and-machine-learning-a087614630fe, Erişim Tarihi: 21.12.2021.