NÖHÜ Müh. Bilim. Derg. / NOHU J. Eng. Sci., 2021; 10(1): 001- 008

Niğde Ömer Halisdemir Üniversitesi Mühendislik Bilimleri Dergisi
Niğde Ömer Halisdemir University Journal of Engineering Sciences
ISSN: 2564-6605
Araştırma / Research

Müşteri davranış tahmini için bir model: Bankacılık sektörü için uygulama

A model for prediction of customer behavior: A case study for banking sector

Kevser Özdem[1,*], M. Ali Akcayol[1]

[1]*Gazi Üniversitesi, Mühendislik Fakültesi, Bilgisayar Mühendisliği Bölümü, 06570, Ankara, Türkiye*

Özet

Son yıllarda kampanya temelli satışlar çok büyük bir hızla artmaya devam etmektedir. Günümüzde birçok sektörde satışlar kampanyaya dayalı olarak gerçekleşmektedir. Bu nedenle kampanya yönetimi oldukça önemli bir konu haline gelmiştir. Literatürde kampanya planlaması ve yönetimi konusunda detaylı bir çalışma bulunmamaktadır. Bu makalede, bankacılık sektöründe kampanya yönetimi ve müşterilerin kampanyaya yönelik ileriye dönük davranışlarının tahmini için bir model geliştirilmiştir. Bankacılık sektörüne özgü geliştirilen birliktelik analizi yapısı kullanılarak bir Portekiz bankasına ait kampanya verilerinden yaygın öğe kümeleri ve birliktelik kuralları oluşturulmuştur. Elde edilen kurallarla kampanyaya katılan müşterilerin ileriye dönük davranışları tahmin edilmiştir. Ayrıca, müşterilerin davranışlarını etkileyen öznitelikler belirlenmiştir. Deneysel sonuçlar, en çok medeni hal ve kredi durumunun müşteri davranışını etkilediğini göstermiştir. Müşterilerin kampanyaya katılıp katılmayacakları geliştirilen model kullanılarak tahmin edilmiştir. Müşterilerin kampanyaya katılım tahmininde %87 oranında başarı sağlanmıştır.

Anahtar kelimeler: Büyük veri analitiği, Veri madenciliği, Birliktelik analizi, Banka pazarlama kampanyası

Abstract

Campaign based sales continue to increase at a very rapid rate in recent years. Today, sales in many sectors are based on the campaign. Therefore, campaign management has become a very important issue. There is no detailed study on campaign planning and management in the literature. In this article, a model is developed for campaign management in the banking sector and prediction of prospective behaviors of customers towards the campaign. Using the association analysis structure developed specifically for the banking sector, frequent itemsets and association rules were created from the campaign data of a Portuguese bank. The prospective behavior of the customers participating in the campaign was estimated with the rules obtained. In addition, attributes that affect the behavior of customers have been identified. Experimental results have shown that marital status and credit status affect customer behavior the most. Using the developed model, a prediction is made on whether the customers will participate in the campaign or not. 87% success was achieved in the prediction of customers' participation in the campaign.

Keywords: Big data analysis, Data mining, Association analysis, Bank marketing campaign

1 Giriş

Günümüzde kampanya temelli satışlar birçok sektörde giderek artan bir öneme sahiptir. Pek çok alanda satışlar kampanyaya dayalı olarak gerçekleşmektedir. Bu değişimin bir sonucu olarak kampanya planlaması ve yönetimi oldukça önemli bir konu haline gelmiştir. Kampanya yönetimi, doğru iletişim tarzını kullanarak doğru müşteriye doğru kanaldan doğru bilgi ve teklifleri sunmaktır [1]. Bankacılık sektörü, bu konuya önem verilen başlıca alanlardan birisidir ve müşteriye uygun planlanmış kampanyalar çok önemlidir. Bu nedenle bankacılık sektöründe müşteri davranışının önceden tahmin edilebilmesi, kurumun pazarlama hedeflerine uygun kampanyalarla müşterilerine mümkün olan en iyi şekilde hitap edebilmesini sağlamaktadır.

Literatürde birliktelik kurallarının farklı alanlardaki kullanımına yönelik çok sayıda çalışma bulunmaktadır, ancak bankacılık sektöründe müşterilerin gelecekteki davranışlarını tahmin etmeye yönelik çalışma bulunmamaktadır.

Hsieh, bir bankadaki kredi kartı müşterilerini yönetmek için entegre bir veri madenciliği ve davranışsal puanlama modeli önermiştir [4]. Farklı müşteri grupları için Apriori algoritması kullanılarak profiller oluşturulmuştur. Bu çalışmada, müşterileri davranışsal olarak tanımlamanın pazarlama stratejisi geliştirmeyi kolaylaştırdığı gösterilmiştir. Birliktelik kuralları, pazarlama alanında yaygın bir şekilde kullanılmaktadır. Literatürde bu alanda yapılmış birçok çalışma bulunmaktadır. Wong ve arkadaşları tarafından gerçekleştirilen çalışmada, doğrudan pazarlamanın etkisinin artırılabilmesi için bir müşteri alt kümesi oluşturulmuştur [5].

Chen ve arkadaşları, müşteri davranışındaki değişiklikleri belirlemek için veri madenciliği yöntemi geliştirmişlerdir [6]. Birliktelik kuralları, müşteri profili ile satın alınan ürün öğeleri arasındaki ilişkiyi tanımlamak için kullanılmıştır. Elde edilen sonuçlara göre müşteri davranışındaki değişiklikler için önerilen yaklaşım, yöneticilerin daha iyi pazarlama stratejileri geliştirmelerine yardımcı olmaktadır.

Moro ve arkadaşları tarafından yapılan çalışmada, pazarlama kampanyalarının veri madenciliği tekniklerinin kullanılarak geliştirilebileceği gösterilmiştir [7]. Bu çalışmada, müşterinin kampanyaya katılım yapmasına göre başarı değerlendirmesi yapan bir model geliştirmişlerdir. Bu modelin, başarıyı etkileyen temel özellikleri belirleyerek, iş gücü, telefon iletişimleri, zaman gibi mevcut kaynakların daha iyi yönetilmesine ve yüksek kaliteli, uygun fiyatlı potansiyel satın alma müşterilerinin seçimine yardımcı olarak kampanya verimliliğini artırdığı gösterilmiştir.

Dongre ve arkadaşları, perakende sektöründe daha iyi uygulamalar için gizli örüntülerin bulunmasının önemli bir problem olduğunu belirtmiştir [8]. Birliktelik kuralları, verisetinden keşfedilen bilgiler kullanılarak tanımlanmıştır. Bu çalışmada kullanılan veriseti üzerinde Apriori

* Sorumlu yazar / Corresponding author, e-posta / e-mail: kevserozdem@gazi.edu.tr (K. Özdem)
Geliş / Recieved: 08.07.2020 Kabul / Accepted: 11.11.2020 Yayımlanma / Published: 15.01.2021
doi: 10.28948/ngumuh.766503

algoritması uygulanmış ve farklı güven değerleri ile ilişkilendirme kuralları oluşturulmuştur. Elde edilen sonuçlara göre perakende sektöründeki verisetlerinde faydalı gizli örüntülerin bulunmasında oluşturulan birliktelik kurallarının faydalı olduğu belirtilmiştir.

Li ve arkadaşları 2015 yılında yaptıkları çalışmada [9], tüketicilerin katılım davranışlarını yakalamak için bir dizi dinamik nitelik kullanımı önermiştir. Çalışmada, pazarlama kampanyalarına yanıt verme ve daha sonra ücretli kullanıcılara dönüşme olasılığı daha yüksek olan kullanıcıların belirlenmesi üzerine çalışılmıştır. Kullanıcının hem görünür hem de gizli davranış kalıplarını yakalayan çeşitli dinamik nitelikler keşfedilmiştir.

Amini ve arkadaşları tarafından yapılan çalışmaya [10] göre, doğrudan pazarlamanın amacı, pazarlama kampanyası mesajlarına yanıt verme olasılığı en yüksek olan doğru müşterileri bulmaktır. Bu müşterileri tahmin etmek için veri madenciliği teknikleri kullanılabilmektedir. Bununla birlikte dengesiz veriler, tahmin işlemine bazı zorluklar getirmektedir. Bu çalışmada, kümeleme ve yetersiz örnekleme ile dengesizlik ortadan kaldırılarak bir sınıflandırma yöntemi önerilmiştir. Deneysel sonuçlar ile önerilen yöntemin, tahmin doğruluğunu artırabildiği gösterilmiştir.

Miguéis ve arkadaşlarının 2017 yılında gerçekleştirdikleri çalışmada [11], bir bankanın gerçek verileri kullanılarak doğrudan pazarlama kampanyalarına verilen yanıtı öngörmek için bir model önerilmiştir. Bu tekniğin bankacılık alanında uygulanması başlangıç aşamasında olduğu belirtilmiştir. Yanıt tahmini için rastgele ormanlardan, sınıf dengesizliği için ise SMOTE ve EasyEnsemble yöntemlerinden yararlanılmıştır.

Parlar 2017 yılındaki çalışmasında [12], müşteri verilerinin sahip olduğu boyut göz önünde bulundurulduğunda, bir insan analist ekibi ile manuel olarak analiz etmenin imkansız olduğunu belirterek, kampanyanın etkinliğini artırmak ve önemli nitelikleri yorumlamak için veri madenciliği tekniklerini kullanmıştır. Önemli nitelikleri seçmek için bilgi kazanımı (information gain) ve Ki-kare yöntemleri göz önünde bulundurulmuştur. Deneysel sonuçlarda, azaltılmış nitelik kümesinin sınıflandırma performansını iyileştirdiğini gösterilmiştir.

Ładyżyński ve arkadaşları, 2019 yılında yaptıkları çalışmada [13], kredi ile ilgilenen müşterileri belirlemek amacıyla rastgele ormanlar ve derin sinir ağları ile sınıflandırmaya dayalı bir sistem önerilmiştir. Ancak, elde edilen bulgular sonucunda, önerilen sistemde iyileştirme ve daha fazla araştırma yapılması gerektiği belirtilmiştir.

2 Müşterilerin ileriye dönük davranış tahmini

Farklı alanlarda başarılı bir şekilde kullanılan veri madenciliği yöntemlerinin başında yaygın öğe kümeleri (frequent itemset) ve birliktelik kuralları gelmektedir.

Veri madenciliği, basit ve açık olmayan, önceden bilinmeyen ve yararlı olan örüntülerin ya da bilginin çok büyük miktarlardaki veriden çıkarılmasıdır. Birliktelik kuralları, veri madenciliğinde ilginç örüntülerin veya ilişkilerin keşfi amacıyla farklı alanlarda yaygın kullanılan bir yöntemdir. En yaygın uygulama alanları olarak market sepeti analizi, tıbbi tanı, protein dizisi analizi, sahtekarlık tespiti ve siber dolandırıcılık tespiti örnek olarak sayılabilir [2, 14-17].

Veritabanlarından kurallar çıkarılması, Agrawal ve Swami tarafından önerildiğinden itibaren farklı alanlardaki çalışmalarda yoğun ilgi görmüştür [3, 14, 15, 18, 19]. Bu kurallar, A ve B nitelik değerleri olmak üzere, $A \longrightarrow B$ şeklinde ifade edilebilmektedir. Bu kurallar, A kümesindeki olayların gerçekleşmiş olması durumunda B kümesindeki olayların gerçekleşme olasılığını incelemektedir.

Birliktelik kurallarının elde edilmesinde ilk aşama, yaygın öğe kümelerinin tümünün oluşturulmasıdır. Bu kümeler, belirlenen eşik destek değerini sağlayacak ölçüde verisetinde bulunan öğelerden oluşmaktadır. Destek sayısı (σ), bir öğeler kümesinin veri kümesinde görülme sıklığı, başka bir deyişle öğeler kümesinin içinde bulunma sayısıdır. Destek oranı (s) ise, öğeler kümesinin içinde bulunduğu hareketlerin toplam hareketlere oranı olarak ifade edilmektedir.

Daha sonra, elde edilen iki ve daha fazla öğeye sahip kümeler kullanılarak belirlenen minimum güven (c) değerinden yüksek güvene sahip kurallar oluşturulmaktadır. $A \longrightarrow B$ kuralı için güven değeri, $A \cup B$ öğeler kümesinin bulunduğu hareketlerin, A öğeler kümesinin bulunduğu hareketlere oranıdır. A ve B verisetine ait öğeler kümesi olmak üzere bu oran Denklem (1) ile ifade edilir.

$$c = \frac{\sigma(A \cup B)}{\sigma(A)} \tag{1}$$

$A \longrightarrow B$ kuralı için güven değeri, A öğesinin bulunduğu herhangi bir küme içerisinde B öğesinin bulunma olasılığı olarak da ifade edilebilmektedir.

Yaygın öğe kümelerinin ve birliktelik kurallarının elde edilebilmesi için gereken yaygın öğe kümelerinin oluşturulmasında çeşitli algoritmalardan yararlanılmaktadır. Apriori, bu algoritmalardan en önemli olanıdır ve oldukça yaygın kullanılmaktadır.

Apriori algoritması, özellikle çok büyük ölçekli verisetleri üzerinde gerçekleştirilen veri madenciliği çalışmalarında yaygın kullanılmaktadır [20, 21-23]. Büyük ölçekli veri setlerinde yüksek ölçeklenebilir, yüksek performans ve kolay uygulanabilir olmasından dolayı başarılı sonuçlar vermektedir. Bu yaklaşım, aşağıdan yukarıya yapılandırma yaparak bir kümenin destek değerinin, boş küme hariç tüm alt kümelerinin destek değerinden büyük olamayacağı kuralını uygular [24].

Birliktelik kuralları farklı uygulamalarda başarılı bir şekilde kullanılmıştır [25- 28]. Ancak, bankacılık sektöründe müşterilerin ileriye dönük davranış tahmini için yapılan çalışma bulunmamaktadır. Bu çalışmada, bankacılık sektöründe kampanya yönetimi amacıyla yaygın öğeler ve birliktelik kuralları probleme özgü bir şekilde uyarlanarak kullanılmıştır. İlk olarak birliktelik kuralları yardımı ile bir tahmin modeli geliştirilmiştir. Daha sonra, bir test kümesi üzerinde bu model kullanılarak ileriye dönük müşterilerin kampanyaya katılıp katılmayacakları tahmin edilmiştir.

Bu çalışmada geliştirilen tahmin modelinde, probleme özgü geliştirilmiş bir birliktelik analizi kullanılmıştır. Bu

analizin gerçekleştirilmesinde hız açısından verimli olması amacıyla yaygın olmayan kümelerin üst kümelerinin de yaygın olamayacağı kuralından yararlanılmıştır.

Öncelikle, verisetinde bulunan yaygın öğeleri oluşturmak amacıyla, bir elemandan başlanarak eşik değeri sağlayacak şekilde sırasıyla daha fazla eleman içeren yaygın öğe kümeleri üretilmiştir. Daha sonra, elde edilen bu yaygın öğe kümelerine geliştirilen birliktelik analizi uygulanarak müşterinin kampanyaya katılımını içeren kurallar oluşturulmuştur.

3 Veriseti

Bu çalışmada kullanılan veriseti, bir Portekiz bankacılık kurumunun doğrudan pazarlama kampanya bilgilerini içermektedir [7]. Pazarlama kampanyalarının en önemli bileşeni telefon görüşmeleridir. Verisetinde, müşteriye ait kişisel bilgiler ve kampanya ile ilgili görüşmeler yer almaktadır ve müşterinin kampanyaya katılıp katılmadığı etiketi belirtilmiştir.

Çalışmada kullanılan verisetinde toplam 45211 kayıt bulunmaktadır ve eksik veri bulunmamaktadır. Bu kayıtların 5289 tanesi kampanyayı kabul eden müşterilere, kalan 39922 tanesi ise kampanyayı reddeden müşterilere aittir. Verisetinden örnek bir kısım Tablo 1'de sunulmuştur.

3.1 Veri önişleme

Veri önişleme aşamasında, 'age', 'balance' ve 'campaign' isimli niteliklere ait değerler gruplandırılmıştır. Veri içerisinde bulunan 'Evet' ve 'Hayır' gibi değerler, özelliğin daha anlaşılabilir olması amacıyla daha açık ifadeler kullanılarak genişletilmiştir. Nitelik çıkarımı (feature extraction) yapılarak çok düşük etkiye sahip ve tekrar eden değerler verisetinden çıkartılmıştır. Veri önişleme aşamasından sonra elde edilen verisetinde bulunan nitelikler ve değerleri Tablo 2'de verilmiştir.

3.2 Eğitim ve test verisetleri

Bu çalışmada veriseti, eğitim verisi (training data) ve test verisi (test data) olarak ikiye bölünmüştür. Verisetinin %70'lik kısmı eğitim verisi olarak, kalan %30'luk kısmı ise test verisi olarak belirlenmiştir. Eğitim ve test verileri rastgele seçilerek oluşturulmuştur. Eğitim verisinde 31647, test verisinde ise 13564 veri kullanılmıştır. Eğitim verisi, kampanya kabul ve ret durumları ile ilgili birliktelik kuralları elde etmek amacıyla, test verisi ise yeni gelecek müşteriler için tahmin amaçlı kullanılmıştır. Eğitim verisi ile elde edilen kurallardan yola çıkılarak geliştirilen model kullanılarak, test verisi üzerinde müşterilerin kampanyaya katılıp katılmayacaklarına ilişkin tahmin yapılmış ve sonuçlar karşılaştırılmıştır.

Tablo 2. Verisetindeki nitelikler ve değerleri

Nitelik	Değerler
Age (Müşterinin yaşı)	(85+, 75-84, 65-74, 55-64, 45-54, 35-44, 25-34, 18-24)
Job (Müşterinin işi)	(admin, unemployed, management, housemaid, entrepreneur, student, blue-collar, self-employed, retired, technician, services)
Marital (Medeni hali)	(married, divorced, single)
Education (Eğitim durumu)	(primary, secondary, tertiary)
Default (Kredi durumu)	(credit, no_credit)
Balance (Bakiyesi-Euro)	(40000+, 20000-40000, 10000-20000, 5000-10000, 1000-5000, 900-1000, 800-900, 700-800, 600-700, 500-600, 400-500, 300-400, 200-300, 100-200, 0-100)
Housing (Konut kredisi)	(housing, no_housing)
Loan (Kişisel kredi)	(loan, no_loan)
Month (Müşteri ile görüşülen ay)	("jan", "feb", ..., "nov", "dec")
Campaign (Kampanya için görüşme sayısı)	(60-69, 50-59, 40-49, 30-39, 20-29, 10-19, 5-9, 1-4)
Y (Kampanya kabul durumu)	(subscribed, deny)

4 Deneysel sonuçlar

Bu çalışmada yapılan deneysel çalışmalar için geliştirilen yazılım, Python programlama dili ile Jupyter Notebook kullanılarak geliştirilmiştir.

Geliştirilen model ile yaygın öğe kümesi seçiminde minimum destek değeri, farklı değerler üzerinde gerçekleştirilen deneyler sonucunda, 0.5 olarak belirlenmiştir. Bu durumda yaygın öğe olarak belirtilen kümelerin, verisetinin yarısında bulunduğu ifade edilmektedir. Seçilen minimum destek değeri ile elde edilen yaygın öğe kümeleri, kampanya kabul durumları {subscribed} için Tablo 3'te ve kampanya ret durumları {deny} için Tablo 4'te verilmiştir

.

Tablo 1. Veriseti alanları ve örnek değerleri

Age	Job	Marital	Education	Default	Balance	Housing	Loan
58	management	married	tertiary	no	2143	yes	no
44	technician	single	secondary	no	29	yes	no
33	entrepreneur	married	secondary	no	2	yes	yes
42	entrepreneur	divorced	tertiary	yes	2	yes	no

Tablo 3. {subscribed} için minimum destek değeri 0.5 olduğunda 1, 2, 3, 4, 5-yaygın öğe kümeleri

Yaygın öğe sayısı	Yaygın öğe kümeleri
1	{1-4}, {married}, {no_credit}, {no_housing}, {no_loan}, {subscribed}
2	{1-4, no_credit}, {1-4, no_housing}, {1-4, no_loan}, {1-4, subscribed}, {married, no_credit}, {married, subscribed}, {no_credit, no_housing}, {no_credit, no_loan}, {no_credit, subscribed}, {no_housing, no_loan}, {no_housing, subscribed}, {no_loan, subscribed}
3	{1-4, no_credit, no_housing}, {1-4, no_credit, no_loan}, {1-4, no_credit, subscribed}, {1-4, no_housing, no_loan}, {1-4, no_housing, subscribed}, {1-4, no_loan, subscribed}, {married, no_credit, subscribed}, {no_credit, no_housing, no_loan}, {no_credit, no_housing, subscribed}, {no_credit, no_loan, subscribed}, {no_housing, no_loan, subscribed}
4	{1-4, no_credit, no_housing, no_loan}, {1-4, no_credit, no_housing, subscribed}, {1-4, no_credit, no_loan, subscribed}, {1-4, no_housing, no_loan, subscribed}, {no_credit, no_housing, no_loan, subscribed}
5	{1-4, no_credit, no_housing, no_loan, subscribed}

Tablo 4. {deny} için minimum destek değeri 0.5 olduğunda 1, 2, 3, 4-yaygın öğe kümeleri

Yaygın öğe sayısı	Yaygın öğe kümeleri
1	{1-4}, {deny}, {housing}, {married}, {no_credit}, {no_loan}, {secondary}
2	{1-4, deny}, {1-4, married}, {1-4, no_credit}, {1-4, no_loan}, {deny, housing}, {deny, married}, {deny, no_credit}, {deny, no_loan}, {housing, no_credit}, {married, no_credit}, {no_credit, no_loan}, {no_credit, secondary}
3	{1-4, deny, no_credit}, {1-4, deny, no_loan}, {1-4, married, no_credit}, {1-4, no_credit, no_loan}, {deny, housing, no_credit}, {deny, married, no_credit}, {deny, no_credit, no_loan}
4	{1-4, deny, no_credit, no_loan}

Kabul durumunda eşiği geçen en fazla beş elemanlı küme elde edilirken ret durumunda en fazla dört elemanlı küme elde edilmiştir. Elde edilen yaygın öğe kümeleri kullanılarak kampanya kabul ve ret durumları ile ilgili minimum güven değeri 0,5 olan kurallar oluşturulmuştur. Böylece yalnızca kabul ve ret ile güçlü bağlantısı bulunan kurallar seçilmiştir. Seçilen birliktelik kuralları için güven, destek ve interest değerleri Tablo 5 ve Tablo 6'da sunulmuştur.

Tablo 5. Ret ile ilgili birliktelik kuralları

Kural	Güven	Destek	Interest
{1-4}–>{deny}	0.88	0.76	-0.01
{housing}–>{deny}	**0.92**	**0.51**	**0.04**
{married}–>{deny}	**0.90**	**0.54**	**0.01**
{no_credit}–>{deny}	0.88	0.87	0.00
{no_loan}–>{deny}	0.87	0.74	-0.01
{1-4, no_credit}–>{deny}	0.87	0.74	-0.01
{1-4, no_loan}–>{deny}	0.87	0.63	0.02
{housing, no_credit}–>{deny}	0.92	0.50	0.04
{married, no_credit}–>{deny}	0.90	0.53	0.01
{no_credit, no_loan}–>{deny}	0.87	0.73	-0.01
{1-4, no_credit, no_loan}–>{deny}	0.87	0.62	-0.02

Elde edilen birliktelik kurallarının güven, destek ve interest değerleri değerlendirildiğinde kampanyaya katılım durumunu en çok müşterinin medeni hali, kredi durumu ve kampanya kapsamında kaç defa arandığının etkilediği görülmektedir. Tablo 5'te koyu olarak belirtilen satırlarda da görülebileceği üzere ev kredisi bulunan müşterilerin %92'si kampanyayı reddetmiştir. Medeni hali evli olan müşterilerin ise %90'ı kampanyayı reddetme eğilimindedir.

Interest değeri esas alınarak yapılan değerlendirmede, Tablo 6'da koyu renkle verilen {1-4, no_credit, no_housing, no_loan} öğe kümesi dikkat çekmektedir. Bu kümenin tüm veriseti içerisindeki destek değeri 0.32 iken kabul eden müşteri verilerindeki destek değeri 0.55 olarak elde edilmiştir. Dolayısıyla, 0.23'lük bir interest değeri ile bu kümede bulunan nitelikler, kabul durumunu destekleyen faktörler olarak belirmektedir. Aynı şekilde koyu olarak verilmiş diğer satırlar da incelendiğinde, 0.23 ve 0.21 farklar ile bu öğe kümelerinin, kabul eden müşteri verileri içerisinde daha yaygın olduğu fark edilmektedir. Yani kampanya kapsamında 1-4 defa aranmış olan ve herhangi bir kredisi bulunmayan müşterilerin kampanyayı kabul etme eğiliminde oldukları görülmektedir.

Müşterinin medeni halinin evli olması ve hali hazırda bir ev kredisinin bulunması ise Tablo 6'da negatif interest değeri ile de anlaşılacağı üzere, kampanyaya ters bir etki yaparak müşteriyi ret durumuna yönlendiren en etkili nitelikler olarak görülmektedir.

Farklı eşik değerleri kullanılarak gerçekleştirilen birliktelik analizleri sonucunda elde edilen bulgular, test verisi üzerine uygulanarak yeni gelecek müşterilere yönelik tahminler gerçekleştirilmiştir. Yapılan tahminler, müşterilerin verdiği kararlar ile karşılaştırıldığında; 0.3 eşik değeri kullanıldığında tahminlerin %82 oranında, 0.5 eşik değeri kullanıldığında %87 oranında, 0.7 eşik değeri kullanıldığında %86 oranında doğru olduğu tespit edilmiştir. Test kümesinde yapılan ileriye dönük tahminin doğruluk değerleri ayrıntılı olarak, 0.3 eşik değeri için Tablo 7'de, 0.5 eşik değeri için Tablo 8'de ve 0.7 eşik değeri için Tablo 9'da verilmiştir.

Tablo 6. Kabul ile ilgili birliktelik kuralları

Kural	Güven	Destek	Interest
{1-4}–>{subscribed}	0.93	0.11	0.06
{married}–>{subscribed}	0.52	0.06	-0.08
{no_credit}–>{subscribed}	0.99	0.12	0.01
{no_housing}–>{subscribed}	0.63	0.07	0.19
{no_loan}–>{subscribed}	0.91	0.11	0.07
{1-4, no_credit}–>{subscribed}	0.92	0.11	0.07
{1-4, no_housing}–>{subscribed}	**0.59**	**0.07**	**0.21**
{1-4, no_loan}–>{subscribed}	0.84	0.10	0.12
{married, no_credit}–>{subscribed}	0.52	0.06	-0.07
{no_credit, no_housing}–>{subscribed}	0.63	0.07	0.19
{no_credit, no_loan}–>{subscribed}	0.90	0.11	0.07
{no_housing, no_loan}–>{subscribed}	**0.59**	**0.07**	**0.21**
{1-4, no_credit, no_housing}->{subscribed}	**0.58**	**0.07**	**0.21**
{1-4, no_credit, no_loan}->{subscribed}	0.84	0.10	0.12
{1-4, no_housing, no_loan}->{subscribed}	**0.55**	**0.06**	**0.23**
{no_credit, no_housing, no_loan}->{subscribed}	**0.59**	**0.07**	**0.21**
{1-4, no_credit, no_housing, no_loan}->{subscribed}	**0.55**	**0.06**	**0.23**

Tablo 7. 0.3 eşik değeri kullanılarak yapılan tahmine ait doğruluk tablosu

	Ret durumu	Kabul durumu	Toplam
Ret tahmini	10 164	648	10 812
Kabul tahmini	1 808	944	2 752
Toplam	11 972	1 592	13 564

Tablo 8. 0.5 eşik değeri kullanılarak yapılan tahmine ait doğruluk tablosu

	Ret durumu	Kabul durumu	Toplam
Ret tahmini	10 610	406	11 016
Kabul tahmini	1 362	1 186	2 548
Toplam	11 972	1 592	13 564

Tablo 9. 0.7 eşik değeri kullanılarak yapılan tahmine ait doğruluk tablosu

	Ret durumu	Kabul durumu	Toplam
Ret tahmini	10 562	520	11 082
Kabul tahmini	1 410	1 072	2 482
Toplam	11 972	1 592	13 564

Tablo 7'ye göre 13564 verinin 10812 tanesi model tarafından ret olarak tahmin edilmiştir. Bunların 10164 tanesi (doğru pozitif) gerçekte de kampanyayı reddeden müşterilerdir. 648 tanesi (yanlış pozitif) ise aslında kampanyayı kabul etmiştir. Gerçekte reddeden 11972 müşteriden 1808 tanesi (yanlış negatif) ise model tarafından tespit edilememiştir.

Tablo 8'e göre 13564 verinin 11016 tanesi model tarafından ret olarak tahmin edilmiştir. Bunların 10610 tanesi gerçekte de kampanyayı reddeden müşterilerdir. 406 tanesi ise aslında kampanyayı kabul etmiştir. Gerçekte reddeden 11972 müşteriden 1362 tanesi ise model tarafından tespit edilememiştir.

Tablo 9'a göre 13564 veriden 11082 tanesi model tarafından ret olarak tahmin edilmiştir. Bunların 10562 tanesi doğru pozitif, 520 tanesi yanlış pozitiftir. Gerçekte kampanyayı reddeden 11972 müşteriden 1410 tanesi ise model tarafından tespit edilemeyerek yanlış negatif kategorisinde belirmiştir.

Bu değerler; literatürde yaygın olarak kullanılmakta olan doğruluk (accuracy), kesinlik (precision), duyarlılık (recall) ve F1-skoru ölçütlerinin hesaplanmasında kullanılmıştır. Doğruluk, yapılan tahminin hangi oranda doğru olduğunu göstermektedir. Bu oran Denklem (2)'de olduğu gibi ifade edilebilmektedir.

$$\text{doğruluk} = \frac{\text{doğru tahmin sayısı}}{\text{toplam tahmin sayısı}} \tag{2}$$

Bu ifade matematiksel olarak, DP doğru pozitif sayısı, DN doğru negatif sayısı, YP yanlış pozitif sayısı ve YN yanlış negatif sayısı olmak üzere, Denklem (3)'te daha detaylı bir şekilde sunulmuştur.

$$\text{doğruluk} = \frac{\text{DP} + \text{DN}}{\text{DP} + \text{DN} + \text{YP} + \text{YN}} \tag{3}$$

Kesinlik ölçütü, ret tahmini yapılanların gerçekte ne kadarının ret olduğunu belirtmektedir. Matematiksel olarak, Denklem (4)'te verildiği üzere, doğru pozitif sayısının toplam pozitif sayısına oranı şeklinde hesaplanabilmektedir.

$$\text{kesinlik} = \frac{\text{DP}}{\text{DP} + \text{YP}} \tag{4}$$

Duyarlılık ölçütü, ret durumlarının ne kadarının ret olarak tahmin edilebildiğini göstermektedir. Bunun için doğru pozitiflerin, doğru pozitifler ve pozitif olması gereken ancak negatif olarak tahmin edilen yanlış negatiflerin toplamına oranı alınmaktadır. Bu oran Denklem (5)'te gösterilmiştir.

$$\text{duyarlılık} = \frac{\text{DP}}{\text{DP} + \text{YN}} \tag{5}$$

F1-skoru ise, Denklem (6)'da ifade edildiği üzere, kesinlik ve duyarlılık değerlerinin harmonik ortalamasını vermektedir.

$$\text{F1 skoru} = 2 \times \frac{\text{kesinlik} \times \text{duyarlılık}}{\text{kesinlik} + \text{duyarlılık}} \tag{6}$$

Geliştirilen model kullanılarak test kümesinde yapılan tahminlere ilişkin ölçütler, farklı eşik değerlere göre Tablo 10'da ve görsel olarak Şekil 1'de verilmiştir.

Tablo 10. Ret tahminine yönelik başarı ölçütleri

Ölçüt	0.3 Eşiği	0.5 Eşiği	0.7 Eşiği
Doğruluk (Accuracy)	0.82	0.87	0.86
Kesinlik (Precision)	0.94	0.96	0.95
Duyarlılık (Recall)	0.85	0.87	0.88
F1-skoru	0.89	0.91	0.91

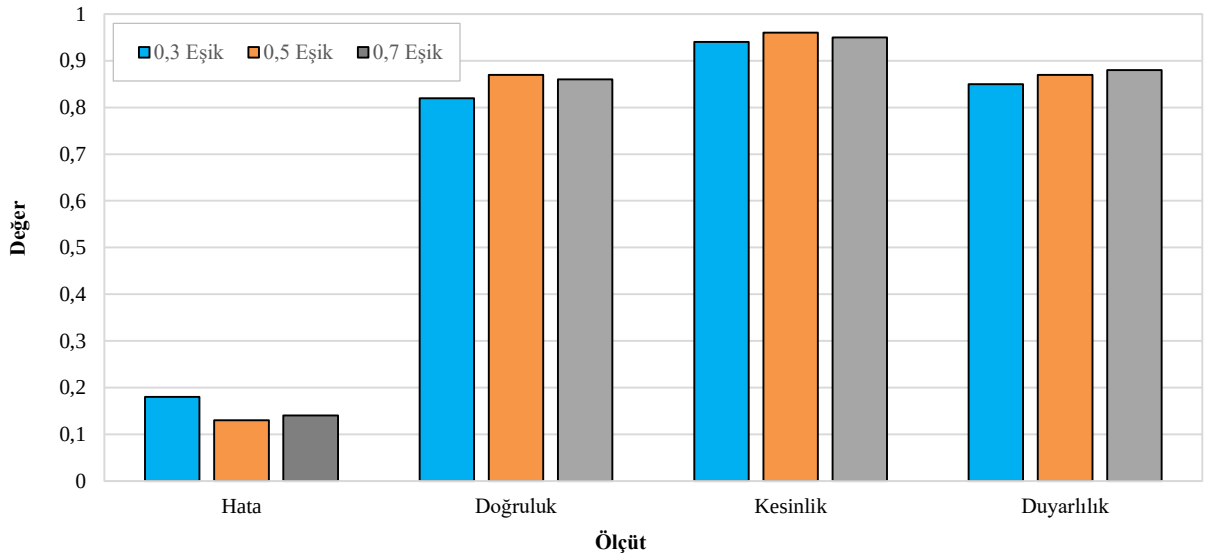
Farklı eşik değerleri ile gerçekleştirilen deneylere ait sonuçlar incelendiğinde, en doğru tahminlerin 0.5 ile yapıldığı gözlemlenmiştir. 0.3 değeri kullanılarak yapılan tahminlerde önemsiz ya da önemi düşük birçok öğe kümesi ve kural da tahmin sürecine dahil edildiğinden yanlış tahminler ortaya çıkmaktadır. 0.7 değeri kullanılarak yapılan tahminlerin sonuçları ise 0.3 ile karşılaştırıldığında daha iyi olmasına rağmen 0.5 ile elde edilen sonuçlar kadar başarılı olamamıştır. Bunun nedeni, aslında ayırt etmede önemli olabilecek kuralların 0.7 ile yapılan deneylerde bulunan fazla kısıtlamadan dolayı sürece dahil edilememesi olarak görülmektedir.

5 Sonuçlar

Bu çalışmada, bankacılık sektöründe kampanya yönetimi için yaygın öğe ve birliktelik kurallarına dayalı bir model geliştirilmiştir. Geliştirilen model ile müşterilerin kampanyaya yönelik ileriye dönük davranışlarının tahmini gerçekleştirilmiştir.

Bu amaçla, veri önişlemeden geçirilen veri setinden farklı minimum destek değerleri kullanılarak yaygın öğe kümeleri ve bu kümelerden yararlanılarak birliktelik kuralları elde edilmiştir. Elde edilen kurallardan yola çıkarak müşterilerin kampanyaya yönelik kabul ve ret davranışlarını tahmin eden bir sistem geliştirilmiştir.

Deneysel sonuçlara göre, müşterinin kampanya kapsamında 1-4 defa aranmış olması ve herhangi bir kredisinin bulunmaması 0.23'lük bir interest değeri ile kabul durumunu desteklemektedir. Aynı şekilde, müşterilerin ev ya da kişisel bir kredisinin bulunmaması, kampanyayı kabul eden müşterilerde %21 oranında daha yaygın bir durumdur. Öte yandan, müşterinin evli olmasının, kampanyayı kabul eden müşteriler içerisinde %8 oranında daha az yaygın olduğu görülmüştür. Tek başına {no_credit} niteliğinin ise hem ret hem de kabul verilerinde yakın bir dağılıma sahip olduğundan, davranış tahmininde bir etkisinin bulunmayacağı gözlemlenmiştir.



Şekil 1. Doğruluk değerleri grafiği

Müşterinin evli olması ve hali hazırda bir ev kredisinin bulunması ise kampanyaya ters bir etki yaparak müşteriyi ret durumuna yönlendirmektedir. Ev kredisi bulunan müşterilerin %92'si kampanyayı reddetmiştir. Medeni hali evli olan müşterilerin ise %90'ı kampanyayı reddetme eğilimindedir. Elde edilen bu sonuçlar, banka tarafından daha sonra oluşturulacak kampanyalarda müşteri davranışlarının tahmin edilmesi veya müşteri profiline uygun kampanya oluşturulması amacıyla kullanılabilecektir.

Bu kapsamda test verisi üzerinde yapılan deneyler sonucunda 0.5 eşik değeri ile eğitilen modelin daha başarılı sonuçlar verdiği görülmüştür. Müşteri davranışı, bu model tarafından %87 oranında, 0.3 eşiği ile oluşturulan model tarafından ise 0.82 oranında bir doğruluk ile tahmin edilebilmiştir. 0.7 ile oluşturulan model ise %86 ile bu orana yakın olsa da gerisinde kalmıştır.

En başarılı model ile yapılan tahminde, kampanyayı reddeden müşterilerin %87'si tespit edilmiştir. Kampanyayı reddedeceği tahmin edilen müşterilerin ise %96'sı gerçekten kampanyayı reddetmiştir. Geliştirilen bu modeli kullanarak bankalar, ileriye dönük müşteri davranış tahmini gerçekleştirebilecektir. Müşteri davranışlarının bu başarı ile önceden tahmin edilebiliyor olması bankanın müşterilerine uygun kampanyalar ile hedefine ulaşmasına yardımcı olacaktır.

Çıkar çatışması

Yazarlar çıkar çatışması olmadığını beyan etmektedir.

Benzerlik oranı (iThenticate): %8

Kaynaklar

[1] H. Hippner, W. Leusser and K. D. Wilde, Campaign management of the fourth generation. Proceedings of 42nd Hawaii International Conference on System Sciences, pp. 1-9, Waikoloa, Hawaii, 5-8 January 2009.

[2] S. T. U. Huq and V. Ravi, Evolutionary multi-objective optimization framework for mining association rules. *arXiv preprint arXiv:2003.09158*, 2020.

[3] R. Agrawal, T. Imieliński, and A. Swami, Mining association rules between sets of items in large databases. ACM SIGMOD international conference on management of data, pp. 207-216, Washington D.C., USA, June 1993.

[4] N. C. Hsieh, An integrated data mining and behavioral scoring model for analyzing bank customers. Expert systems with applications, 27 (4), 623-633, 2004. https://doi.org/10.1016/j.eswa.2004.06.007.

[5] K. W. Wong, S. Zhou, Q. Yang, and J. M. S. Yeung, Mining customer value: From association rules to direct marketing. Data Mining and Knowledge Discovery, 11 (1), 57-79, 2005. https://doi.org/10.1007/s10618-005-1355-x.

[6] M. C. Chen, A. L. Chiu, and H. H. Chang, Mining changes in customer behavior in retail marketing. Expert Systems with Applications, 28 (4), 773-781, 2005. https://doi.org/10.1016/j.eswa.2004.12.033.

[7] S. Moro, R. Laureano, and P. Cortez, Using data mining for bank direct marketing: An application of the crisp-dm methodology. Proceedings of European Simulation and Modelling Conference-ESM '2011, pp. 117-121, Guimaraes, Portugal, 24-26 October 2011.

[8] J. Dongre, G. L. Prajapati, and S. V. Tokekar, The role of Apriori algorithm for finding the association rules in data mining. Proceedings of 2014 International Conference on Issues and Challenges in Intelligent Computing Techniques (ICICT), pp. 657-660, Ghaziabad, India, 7-8 February 2014.

[9] Y. Li, P. Murali, N. Shao, and A. Sheopuri, Applying data mining techniques to direct marketing: challenges and solutions. Proceedings of 2015 IEEE International Conference on Data Mining Workshop (ICDMW) '11, 2015, pp. 319-327, Atlantic City, NJ, USA, 14-17 November 2015.

[10] M. Amini, J. Rezaeenour, and E. Hadavandi, A cluster-based data balancing ensemble classifier for response modeling in bank direct marketing. International Journal of Computational Intelligence and Applications, 14 (4), 1550022, 2015. https://doi.org/10.1142/S1469026815500224.

[11] V. L. Miguéis, A. S. Camanho, and J. Borges, Predicting direct marketing response in banking: comparison of class imbalance methods. Service Business, 11 (4), 831-849, 2017. https://doi.org/10.1007/s11628-016-0332-3.

[12] T. Parlar, Using data mining techniques for detecting the important features of the bank direct marketing data. International Journal of Economics and Financial Issues, 7 (2), 692, 2017.

[13] P. Ładyżyński, K. Żbikowski, and P. Gawrysiak, Direct marketing campaigns in retail banking with the use of deep learning and random forests. Expert Systems with Applications, 134, 28-35, 2019. https://doi.org/10.1016/j.eswa.2019.05.020.

[14] M. A. Valle, G. A. Ruz, and R. Morrás, Market basket analysis: Complementing association rules with minimum spanning trees. Expert Systems with Applications, 97, 146-162, 2018. https://doi.org/10.1016/j.eswa.2017.12.028.

[15] G. Agapito, P. H. Guzzi, and M Cannataro, Parallel extraction of association rules from genomics data. Applied Mathematics and Computation, 350, 434-446, 2019. https://doi.org/10.1016/j.amc.2017.09.026.

[16] A. Verma, A. Taneja, and A. Arora, Fraud detection and frequent pattern matching in insurance claims using data mining techniques. Tenth International Conference on Contemporary Computing (IC3), pp. 1-7, Noida, India, 10-12 August 2017.

[17] K. C. Lekha, and S. Prakasam, Data mining techniques in detecting and predicting cyber crimes in banking sector. International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS), pp. 1639-1643, Chennai, India, 1-2 August 2017.

[18] C. C. Shen, L. Y. Hu, and Y. H. Hu, Comorbidity study of borderline personality disorder: applying association rule mining to the Taiwan national health insurance

research database. BMC Medical Informatics and Decision Making, 17 (1), 8, 2017. https://doi.org/10.1186/s12911-016-0405-1.

[19] H. Changhai, and H. Shenping, Factors correlation mining on maritime accidents database using association rule learning algorithm. Cluster Computing, 22 (2), 4551-4559, 2019. https://doi.org/10.1007/s10586-018-2089-z.

[20] R. Agrawal, and R. Srikant, Fast algorithms for mining association rules. Proceedings of 20th int. conf. very large data bases, VLDB, pp. 487-499, Santiago de Chile, Chile, 12-15 September 1994.

[21] S. Rathee, and A. Kashyap, Adaptive-Miner: An efficient distributed association rule mining algorithm on Spark. Journal of Big Data, 5 (1), 6, 2018. https://doi.org/10.1186/s40537-018-0112-0.

[22] S. Biswas, N. Biswas, and K. C. Mondal, Parallel Apriori based distributed association rule mining: A comprehensive survey. Fourth International Conference on Research in Computational Intelligence and Communication Networks (ICRCICN), pp. 202-207, Kolkata, India, 22-23 November 2018.

[23] E. Stamoulakatou, A. Gulino, and P. Pinoli, DLA: A distributed, location-based and Apriori-based algorithm for biological sequence pattern mining. IEEE International Conference on Big Data (Big Data), pp. 1121-1126, Seattle, WA, USA, 10-13 December 2018.

[24] B. Liu, Web data mining: exploring hyperlinks, contents, and usage data. Leipzig: Springer Science & Business Media, 2007. https://doi.org/10.1007/978-3-540-37882-2

[25] H. Yang, C. Rudin, and M. Seltzer, Scalable Bayesian rule lists. Proceedings of International Conference on Machine Learning, pp. 3921-3930, Sydney, Australia, 6-11 August 2017.

[26] T. Wang, Multi-value rule sets for interpretable classification with feature-efficient representations. Proceedings of Advances in Neural Information Processing Systems, pp. 10835-10845, Montreal, Canada, 2-8 December 2018.

[27] C. Song, and T. Ge, Discovering and managing quantitative association rules. Proceedings of the 22nd ACM international conference on Information & Knowledge Management, pp. 2429-2434, San Francisco, California, USA, October 2013.

[28] F. Mendoza Palechor, A. Carrascal Oviedo, and E. De la Hoz, Association rules implementation for affinity analysis between elements composing multimedia objects. Journal of Theoretical and Applied Information Technology, 97 (6), 1764-1774, 2019. Available: http://hdl.handle.net/11323/5262.