

Doğrudan Pazarlama Amaçlı Hedef Kitle Analizi Target Audience Selection for Direct Marketing

Sinan Keçeci¹, Eyüp Erkan Özbek¹, Mustafa Sertaç Türkel¹, Remzi Düzağaç¹
Veri Bilimi ve Analitik Bölümü¹
ETSTUR¹
İstanbul, Türkiye¹
sinan.kececi, eyup.ozbek, sertac.turkel, remzi.duzagac@etstur.com¹
Olca Taner Yıldız²
Işık Üniversitesi Bilgisayar Mühendisliği Bölümü²
olca.yildiz@isikun.edu.tr²

Özetçe —Doğrudan pazarlama, uygun ürünleri uygun kişilerle en kısa yoldan buluşturma sürecidir. Son yılların en popüler pazarlama yaklaşımlarından birisidir. Bu çalışmada turizm sektörüne ait isimsizleştirilmiş bir veri tabanını kullandık. Bir otel zinciri için yapılan kampanya kapsamında veri madenciliği tekniklerini uygulayarak hedef kitle seçimi yaptık. Çalışmada birçok makine öğrenmesi yöntemini denedik. Sonuç olarak; geçmişte yapılan ve herhangi bir makine öğrenmesi yöntemi kullanılmadan hazırlanan kampanya sonuçlarına göre daha iyi sonuçlar elde ederken benzer analizlerde kullanılacak bir altyapı oluşturmuş olduk.

Anahtar Kelimeler—*Makine Öğrenmesi, Doğrudan Pazarlama, Veri Madenciliği, Öznitelik Mühendisliği*

Abstract—Direct marketing, which is one of the most popular marketing approaches in the recent years, can be defined as the shortest and most accurate way of matching suitable products with the appropriate people. In this paper, we used an anonymous tourism sector database. We applied data mining techniques for target audience selection for a hotel campaign. As a result; we achieved better results than the results of previous campaigns without the use of any machine learning methods and we created an infrastructure that can be used for similar analyses.

Keywords—*Machine Learning, Direct Marketing, Data Mining, Feature Engineering*

I. GİRİŞ

Kurumlar için geçmişte veriyi depolamak çok önemli olmasına rağmen günümüzde sahip olunan veriyi verimli bir şekilde yönetmek ve işleyebilmek çok daha önemli hale gelmiştir. Kurumlar teknolojinin verdiği imkân dahilinde çok yüksek hacimde operasyonel veriye sahiptirler ve yakın geçmişe kadar potansiyel tanıtım yapılabilecek kişilere ulaşabilmek için sadece televizyon, radyo gibi kitlesel yayın yapan araçlar ile toplu pazarlama yöntemlerini kullanmışlardır. Gelişen teknoloji, veriyi etkin bir şekilde analiz edebilmeye

olanak sağlamıştır. Bununla birlikte, artan rekabet koşulları ve maliyetler şirketleri pazarlama stratejilerini belirlerken daha verimli ve hızlı dönüş alabilecekleri, doğru kitleyle doğru zamanda iletişim kurmaya yardımcı olacak analizler yapmaya yöneltmiştir.

Bu çalışmadaki hedefimiz toplu pazarlama yöntemleri yerine doğrudan pazarlama yöntemlerini en verimli şekilde kullanarak ürünleri uygun kitleyle buluşturmaktır. Doğrudan pazarlama, iletişim araçları kullanılarak bireylere doğrudan ulaşılması ve ürün tanıtımı yapılması ile gerçekleştirilmektedir. Doğrudan pazarlama kanallarıyla ulaşılacak kitlenin seçimi için öncelikle sunulacak kampanya ve tanıtılacak ürünü satın alan kişilerin özellikleri belirlenmelidir. Bu ayırt edici özelliklerin belirlenmesi için veri madenciliği yöntemleri kullanılmalıdır. Veri madenciliği metodları ile yapılan işlemler sonrasında oluşan modeller bize ürünlerin tanıtımını yapabileceğimiz kitlenin özelliklerini daha iyi tanımlamamıza ve bu kişileri bulmamıza olanak sağlar.

Biz bu bildiriye hedef kitlenin bulunması için 2 sınıflı bir sınıflandırma yöntemi önermekteyiz. Bu sınıflar için seçilen örnekler söz konusu kampanyalar için hedef ve hedef olmayan kişilerin özelliklerini göstermekte ve sınıflandırma yöntemleri kullanılarak veri tabanından iletişim kurulacak kişiler seçilmektedir.

Çalışmamızda turizm sektöründe faaliyetlerini sürdüren bir şirketin verisi kullanılmıştır. Veri niteliğinin nasıl geliştirildiği, özniteliklerin nasıl hazırlandığı ve oluşturulan modellerin gerçek hayattaki başarısı bu bildiri kapsamında anlatılacaktır. Bu bildiriye konu olan yöntemler her ne kadar turizm sektöründeki bir veri sistemine uygulanmış olsa da, verilerin oluşturulma şekli ve oluşabilecek sorunlara getirilen çözümler diğer sektörlerdeki veri sistemlerine de uygulanabilir.

Makalenin II. bölümünde ilgili literatür incelemesine, III. bölümünde kullanılan verinin niteliğine ve seçilen özniteliklere, IV. bölümünde uygulanan farklı makine öğrenimi algoritmalarının test seti üzerindeki başarısına, V. bölümünde

uygulanan modelin gerçek hayattaki başarısına değinilmiştir.

II. LİTERATÜR İNCELEMESİ

Turizm sektörü için yaptığımız bu çalışmaya, farklı sektörlerdeki çalışmalar referans olmuştur. Ayetiran ve diğerleri makalelerinde [1] bankacılık sektöründen bir veri kümesi üzerinde çalışma yapmışlardır ve bankacılık sektöründe makine öğrenmesi yöntemlerinin hedef kitle seçimi sırasında ne kadar verimli kullanılabileceğini araştırmışlardır. Çalışmada naive Bayes sınıflandırma algoritması kullanılmıştır ve makine öğrenmesi yöntemlerinin etkin pazarlama stratejileri geliştirilmesine yardımcı olduğu gözlemlenmiştir.

Bir başka hedef kitle seçimi çalışmasında, Ling ve diğerleri [2] üç farklı firmaya ait veri kullanmışlardır. Bunlar sırasıyla; bankacılık, sigortacılık ve farklı şirketler için bonus kampanyaları yürüten bir firmadır. Çalışmada bu üç veri kümesi için karşılaşılan problemler ve çözümlerine ilişkin bulgular paylaşılmıştır ve bu veri kümeleri üzerinde makine öğrenmesi algoritmaları çalıştırarak sonuçlar gözlemlenmiştir. Çalışmada hedef kitleyi bulmak için naive Bayes ve karar ağacı algoritmaları kullanılmıştır. Her iki algoritmada da elde edilen bulgular, yapılan bu çalışmanın daha etkin pazarlama yapılabilmesi açısından gerekli olduğunu göstermiştir.

Doğrudan pazarlama ile ilgili bankacılık sektöründeki bir başka çalışma ise yapay sinir ağları ve karar ağaçlarını birlikte kullanan Elsalamony ve diğerlerinin [3] hedef kitle seçimi çalışmasıdır. Yapılan çalışmada bankacılık sektörüne ait California Üniversitesi tarafından hazırlanmış bir veri kümesi kullanılmıştır. Bu veri kümesi üzerinde çok katmanlı sinir ağları ve Ross Quinlan'ın yeni karar ağacı modeli hedef müşterileri tespit etmek için kullanılmıştır. Bu çalışmanın sonucunda karar ağacı modelinin daha iyi bir performans gösterdiği gözlemlenmiştir.

Sing'oei ve diğerleri [4] ise makalelerinde veri madenciliği ile birlikte hedef kitle seçimine yaklaşımlarını bir çatı çerçevesinde sunmuşlardır ve bankacılık sektöründe kullanılan bir veri kümesi üzerinde karar ağacı algoritmasını uygulayıp sonuçları gözlemlenmişlerdir. Bildiriye göre bu algoritmalar kullanılarak yapılan çalışmalar etkin pazarlama yapılabilmesi açısından önem arz etmektedir.

Wisaeng [5] makalesinde; bankacılık sektöründe Doğrudan pazarlama amaçlı yapılmış bir hedef kitle seçimi uygulamasını anlatmıştır. Çalışmasında karar ağacı yöntemleri, radyal bazlı fonksiyon ağları ve destek vektör makinesi gibi makine öğrenmesi yöntemlerini denemiştir. Destek vektör makinesi algoritması ile en iyi sonuca ulaşmıştır.

Das [6] geçmişteki satın alma hareketlerini kullanarak, hangi müşterilerin yeni ürünleri almaya daha yatkın olacağını belirlemeye yönelik yaptığı çalışmasında naive Bayes, k en yakın komşu, destek vektör makinesi algoritmalarını uygulamıştır. En iyi sonucu naive Bayes algoritması ile almıştır.

III. VERİ NİTELİĞİ

Hedef kitle seçiminin verimli bir şekilde yapılabilmesi için üzerinde çalışılan veri tabanının, kişilerin satın alma özelliklerini çok iyi temsil edecek özniteliklere sahip olması gerekmektedir. Bildirinin bu bölümünde yaptığımız analizlerde kullanılan öznitelikleri nasıl oluşturduğumuzu anlatmaktayız.

A. Genel Öznitelikler

Üzerinde çalıştığımız firmanın veri tabanındaki ürünler otel ve tur rezervasyonlarından oluşmaktadır. Bu ürünleri satın alan kişilerin tanımlanması için demografik bilgilerinin dışında satın alma alışkanlıklarını daha iyi tanımlayacak oransal değerler öznitelik olarak belirlenmiştir. Aynı zamanda kişilerin ürün satın alma zamanları, öznitelikler belirlenirken kullanılmıştır. Kullanılan özniteliklerinin açıklamaları TABLO I' de verilmiştir.

B. Güncellik, Frekans ve Tutar Özniteliği

Analizlerimizde öznitelik olarak kullanılan bir başka kavram ise müşterinin güncellik, frekans ve harcadığı tutar bilgileri kullanılarak kişilerin şirkete kattığı değere göre hesaplanan müşteri puanıdır. Bu değerın çıkışı Pareto ilkesine dayanmaktadır. Pareto ilkesine göre etkilerin ortalamada yüzde 80'i etkenlerin yüzde 20'sinden kaynaklanmaktadır. Bu kural iş dünyasına yönelik olarak yorumlanırsa bir şirket gelirinin yüzde 80'ini müşterilerinin yüzde 20'sinden elde etmektedir. Müşteri değeri hesaplanması için 3 müşteri özelliği kullanılmaktadır. Bu özellikler aşağıdaki verilmiştir:

- Güncellik: Müşterinin ay veya yıl cinsinden en son ürün satın almasından günümüze kadar geçen zaman

- Frekans: Müşterilerin belirli bir süre içinde satın aldığı ürün sayısı

- Tutar: Müşterinin harcadığı toplam tutar

Bu özellikler için her biri, müşteri sayısının %20'sini içerecek 5 ayrı değer aralığı belirlenir. Bu değer aralıkları en önemliden en önemsize 5'ten 1'e doğru numaralandırılır. Bu puanlar güncellik, frekans ve toplam harcama tutarı olmak üzere sırasıyla yan yana yerleştirilir ve 3 basamaklı müşteri değeri elde edilir. Bu şekilde şirket için en değerli müşteri grubu belirlenen her özelliğe ilk yüzde 20'lik kısma giren 555 puanlı, en değersiz müşteri grubu ise belirlenen her özelliğe son yüzde 20'lik kısma giren 111 grubudur. [7]

C. Eğitim Kümesi Seçimi

Çalışmamızda, kampanyası yapılacak ürünü satın alma potansiyeli olan hedef kitle ve ürünü almayacağını düşündüğümüz hedef olmayan kitle şeklinde sınıflara sahip 2 sınıflı bir model kullanılmıştır.

Yapılan çalışmalara örnek olarak 2017 senesinde bir otel grubuna yönelik uygulanan pazarlama kampanyası için hazırlanan analiz anlatılacaktır. Bu analizin seçilme nedeni daha önceki yıllarda benzer kampanyaların yapılmış olması ve sonuçların Doğrudan olarak karşılaştırılabilmesidir.

Eğitim kümelerinin hangi koşullar ile oluşturulduğu Tablo II' de verilmiştir. Analizi yapılacak ürünün eğitim setini oluştururken karşılaşılabilecek sorunlardan birisi dengesiz örnek sayılarına sahip sınıflardır. Bu sorunun nedeni birçok ürün çeşidine sahip bir şirkette bir ürünü alanların genele göre küçük bir yüzdeyi oluşturmasıdır. Bu sorunu gidermek ve eğitim kümesini dengelemek için satır bazında alt örnekleme kullanılmıştır. Yaptığımız analizde alt örnekleme sonucunda her biri 300.000 örnekten oluşan iki sınıflı bir eğitim kümesi elde edilmiştir.

TABLO I: VERİ ÖZNETELİKLERİ VE AÇIKLAMALARI

| Özniteliklerin Genel Açıklaması | İçerdiği Veritabanı Alanları |
|---|--|
| Demografik bilgiler (İsimsizleştirilmiş) | Yaşı, yaşadığı bölge, mesleki bilgiler gibi karakteristikler |
| Toplam ürün satın alma sayısı | Toplam rezervasyon sayısı |
| Bir seferde ortalama harcanan para | Yapılan rezervasyonların ortalama tutarı |
| Bir seferde harcanan maksimum ve minimum tutarlar | Kişinin yaptığı rezervasyonlar içinde maksimum ve minimum rezervasyon tutarları |
| Ürünlerin detayları ile ilgili oranlar ve minimum, maksimum değerler | Bir kişinin rezervasyonlarının ortalama konaklama süresi, konaklama süresinde kişi başı harcanan ortalama para ve bunlarla ilgili minimum, maksimum değerler |
| Grup şeklinde alınan bir ürün ise ürünü alan grubu tanımlayan oranlar | Rezervasyonlara katılan ortalama yetişkin ve çocuk sayısı |
| Satın alınan ürün gruplarının sayıları ve satın alma oranları | Tatil köyü, yurtdışı turları, kültür turları gibi 12 adet ürün grubu satışlarının sayısı ve toplam satılan ürünlere oranı |
| Aktif alışveriş yaptığı süre | Kişilerin rezervasyon yaptırdığı sene sayısı |
| Aylara göre satın alma oranı | Dönemsel davranışları belirlemek için kişilerin her ay hangi oranda alışveriş yaptığı bilgisi |
| Ürünlerin satın alınması arasında geçen ortalama süre | İki rezervasyon arasında geçen ortalama süre |
| Kampanyaların tercihi ile ilgili oranlar | Kişilerin erken rezervasyon kampanyalarını kullanma oranı |
| Ürünün satın alınması ve kullanılması arasında geçen ortalama süre | Kişilerin rezervasyon yaptırdıktan sonra ortalama ne kadar süre sonra tatile gittiği bilgisi |
| Satış kanallarının tercih edilme oranları | İnternet sitesi, turizm acenteleri ve çağrı merkezi gibi satın alma kanallarından alınan rezervasyon sayısının toplam rezervasyon sayısına oranı |

Bu bildiriye sonuçları paylaşılacak analizlerin hepsinde, tanıtımı yapılmak istenen ürünleri daha önce satın alan kişiler bulunmaktadır. Eğer yeni bir ürün için benzer bir çalışma yapılacaksa pilot bir müşteri grubu içerisinde ürünü satın alacaklara bakılmalı ve daha sonra veri madenciliği teknikleri ile bu grubun özelliklerine benzer kişiler genel veri tabanından seçilmelidir.

TABLO II: EĞİTİM KÜMESİ SEÇİMLERİ

| Analiz İsmi | Eğitim Kümesi Sınıfı | Eğitim Kümesi Sınıfının Açıklaması |
|---|----------------------|---|
| Otel Kampanyası için Hedef Kitle Seçimi | Hedef Kitle | Analizi yapılacak otel ve benzer özellikteki otellerden daha önce rezervasyon yaptırmış kişiler |
| | Hedef Olmayan Kitle | Analizi yapılacak otel ve benzer özellikteki otellerden daha önce rezervasyon yaptırmamış fakat başka otellerden birden fazla kez rezervasyon yaptırmış kişiler |

IV. HEDEF KİTLE SINIFLANDIRMA SONUÇLARI

Hedef ve hedef olmayan kitlelerin sınıflandırılması için tercih edilen algoritmalar bu konuda daha önce yapılan çalışmalarda olumlu sonuçlar verdiği görülmüş rassal ormanlar, yapay sinir ağları ve naive Bayes'tir. Sınıflandırma yapılmadan önce her öznetelikte en alt ve en üst yüzde 1'lik dilime giren aykırı değerler satır bazında çıkarılmış ve öznetelikler en küçük değer 0, en büyük değer 1 olacak şekilde 0 ile 1 arasına ölçeklenmiştir.

Algoritmaların performans sonuçları TABLO III'te verilmiştir.

TABLO III: ALGORİTMA PERFORMANS SONUÇLARI

| Algoritma | Doğruluk | Kesinlik | Duyarlılık |
|--------------------|--------------|--------------|--------------|
| Rassal Ormanlar | 82,64 ± 0,20 | 83,84 ± 0,28 | 80,87 ± 0,18 |
| Yapay Sinir Ağları | 81,67 ± 0,19 | 81,33 ± 0,51 | 82,24 ± 1,05 |
| Naive Bayes | 72,31 ± 0,24 | 71,73 ± 0,88 | 73,71 ± 1,70 |

TABLO III'te görülen test sonuçları eğitim kümesinin 5 katmanlı çapraz geçeri ile elde edilmiştir. Doğruluk, kesinlik ve duyarlılık için ortalama başarı oranları ve ortalama standart sapmalar verilmiştir. Algoritmalar için Python'ın scikit-learn modülü kullanılmıştır. Algoritma parametrelerinin optimizasyonu için her bir parametre belirli değer aralıklarında denenmiş ve algoritmanın performansına bakılmıştır. Rassal ormanlar için kullanılacak ağaç sayısına 100 ile 600 arasında bakılmış ve bu parametre 400 olarak belirlenmiştir. Yine rassal ormanlarda her bir ayırmada kaç öznetelik inceleneceği sayısı 2 ile 25 arasında incelenmiş ve bu parametre 5 olarak belirlenmiştir. Yapay sinir ağlarında gizli katman sayısı 2 ile 10 arasında incelenmiş ve katman sayısı için 4 seçilmiştir.

Doğruluk ve kesinlik oranlarındaki üstünlüğünden dolayı analizin uygulama kısmı için rassal ormanlar algoritması seçilmiştir.

V. UYGULAMANIN GERÇEK HAYATTAKİ BAŞARISI

Eğitim kümesi üzerinde rassal ormanlar algoritması kullanılarak hazırlanan makine öğrenimi modeli daha sonra veri tabanının tamamına uygulanmış ve hedef kitle belirlenmiştir. Bu seçilen kişilere Doğrudan pazarlama ile ulaşılmış ve kampanya gerçekleştirilmiştir. Çalışmamız kapsamında makine öğrenimi ile hedef kitle seçimi yapılmış bu kampanyanın, makine öğrenimi yapılmadan gerçekleştirilen kampanyalara göre ne kadar başarılı olduğunu karşılaştırmak için 2015 ve 2016 senelerine ait benzer kampanyaların geri dönüş oranlarına bakılmıştır. 2015 ve 2016 yıllarında yapılan benzer kampanyaların her biri için kampanya başlangıcından itibaren 3 ay içerisinde alışveriş yapan kişi sayısının toplam kampanya yapılan kişi sayısına oranı belirlenmiştir. Son olarak bu oranların ortalaması alınmıştır. Bu geri dönüş oranları ile bu bildiriye anlatılmış olan yöntem ile hazırlanmış kampanyanın 3 ay içerisindeki geri dönüş oranı karşılaştırılmıştır. Bu karşılaştırma TABLO IV'te verilmiştir.

TABLO IV: YAPILAN ÇALIŞMANIN GERİ DÖNÜŞÜNÜN ÖNCEKİ BENZER ÇALIŞMALARLA KARŞILAŞTIRILMASI

| 2015-2016 Yıllarında Yapılan Kampanyanın Geri Dönüş Oranı | | 2017'de Makine Öğrenimiyle Yapılan Kampanyanın Geri Dönüş Oranı | 2017'deki Geri Dönüşün İyileşme Oranı |
|---|-------|---|---------------------------------------|
| 2015 | 0,140 | 0,24 | 1,71 |
| 2016 | 0,108 | 0,24 | 2,22 |
| Ortalama | 0,124 | 0,24 | 1,94 |

VI. SONUÇ

Bu çalışmada turizm sektöründeki bir firmanın kampanyası için bu firmanın veri tabanı üzerinde uyguladığımız veri madenciliği yöntemlerini ve aldığımız sonuçları paylaştık.

Çalışmada veri madenciliği ile yapılan doğrudan pazarlamanın ne kadar etkili olabileceğini gerçek hayattaki sonuçlarla görmüş olduk. Yaptığımız çalışma sonucunda iletişime geçilen kişilerden yüzde 24 oranında olumlu geri dönüş elde ettik. Herhangi bir veri madenciliği yönteminin kullanılmadığı ve hedef kitlenin veri tabanından belli özelliklerdeki kişiler seçilerek yapıldığı 2015 ve 2016 yıllarındaki kampanyaların sonuçlarına göre yaklaşık yüzde 94'lük bir artış sağladık.

Bu çalışmayla birlikte gelecek çalışmalarımız için bir alt yapı ve bilgi birikimi elde etmiş olduk. Kullandığımız özniteliklerimize tercih edilen ürünlerin detayları, kişilerin diğer özellikleri gibi yeni öznitelikler ekleyerek ve yeni yöntemler kullanarak benzer analizlerde elde ettiğimiz sonuçları daha da geliştirmeye çalışacağız.

KAYNAKLAR

- [1] E. F. Ayetiran, A. B. Adeyemo, "A Data Mining-Based Response Model for Target Selection in Direct Marketing," International Journal of Information Technology and Computer Science, vol. 4, no. 1, pp. 9–18, Mar. 2012
- [2] C.X. Ling, C. Li, "Data mining for direct marketing: Problems and solutions," In KDD, volume 98, pages 73–79, 1998.
- [3] H. Elsalamony, A. Essayed, "Bank Direct Marketing Based on Neural Network," International Journal of Engineering and Advanced Technology (IJEAT) ISSN: 2249 – 8958, Volume-2, Issue-6, August 2013.
- [4] L. Sing'oei, J. Wang, "Data Mining Framework For Direct Marketing: A Case Study of Bank Marketing," IJCSI International Journal of Computer Science Issues, Vol. 10, Issue 2, No 2, March 2013
- [5] K. Wisaeng, "A Comparison of Different Classification Techniques for Bank Direct Marketing," International Journal of Soft Computing and Engineering (IJSCE) ISSN: 2231-2307, Volume-3, Issue-4, September 2013.
- [6] T. K. Das, "A customer classification prediction model based on machine learning techniques," 2015 International Conference on Applied and Theoretical Computing and Communication Technology (iCATccT), 2015.
- [7] Derya Birant, "Data Mining Using RFM Analysis" Knowledge- Oriented Applications in Data Mining, InTech, 2011.