

İLERİ MÜHENDİSLİK ÇALIŞMALARI VE TEKNOLOJİLERİ DERGİSİ

Müşteri Kayıp Analizi İçin Sınıflandırma Algoritmalarının Karşılaştırılması

Alper ÇİÇEK*¹ , Yelda ARSLAN*² 

*İstanbul Aydın Üniversitesi, Fen Bilimleri Enstitüsü, Bilgisayar Mühendisliği Ana Bilim Dalı, İstanbul, 34295, Türkiye

Araştırma Makalesi, Geliş Tarihi: 06.07.2020, Kabul Tarihi: 22.07.2020

Özet

Bu çalışmada kaggle.com üzerinden yayımlanan toplam iki veri seti ile çalışılmıştır. İlk veri seti bir banka müşterilerine, ikinci veri seti ise telefon operatörü müşterilerine aittir. Kullanılan iki veri seti için de çeşitli nitelikler incelenerek müşteri kayıp analizi yapılmıştır. Çalışmada Logistic Regression, Naive Bayes, Desicion Tree, K-NN, SVM ve LDA sınıflandırma modelleri kullanılmıştır. İncelemede toplam 13.000 müşteri bilgisi üzerinden belirlenen yöntemler ile çalışma yapılmıştır. Algoritmanın uygulanması kolay ve literatür çalışması fazla olduğu Python Programlama Dili seçilmiştir. Kullanılan sınıflandırma yöntemleri her iki veri seti için de uygulanmış ve doğruluk oranları karşılaştırılmıştır. En kararlı sonucu karar ağaçları algoritması olduğu gözlemlenmiştir. Tüm sonuçların %70'ten daha fazla bir doğruluk oranı vermesi başarılı bir çalışma yapıldığını ortaya çıkarmaktadır.

Anahtar Kelimeler: Müşteri Kayıp Analizi, Logistic Regression, Naive Bayes, Desicion Tree, K-NN.

Comparing Classification Algorithms of Customer Loss Analysis

Abstract

Total of two data sets which is published from kaggle.com has been used in this study. The first data set belongs bank customers, and the second data set contains telephone operator customers. Customer loss analysis was conducted by examining various attributes with using the two data sets. In this study, Logistic Regression, Naive Bayes, Desicion Tree, K-NN, SVM and LDA classification modeling are used. The methods which are determined from 13.000 customer information are used in this research. Python Programming Language, which is easy to implement and has many literature studies, has been chosen. The classification methods used were applied for both data sets and their accuracy rates were compared. It has been observed that the most decisive result is the decision trees algorithm. More than 70% of all results give an accuracy rate, a successful study is revealed.

Keywords: Customer Loss Analysis, Logistic Regression, Naive Bayes, Desicion Tree, K-NN.

¹Sorumlu yazar alpercicek@msn.com, ²arslanyelda@hotmail.com

1. GİRİŞ

Günümüzde firmalar için önemli bir konu haline gelen müşteriye elinde tutma, geri kazanma gibi konular popüler haldedir (Accardi ve Petersen, 2011). Artan rekabet ve pazarın odak noktasının müşteri olması, ilgiyi bu alana çekmiştir. Böylelikle her geçen gün müşteriye firmaya bağlı tutma veya geri kazanma gibi yeni geliştirilen ve üzerinde durulan noktalardır. Söz konusu firmanın devamlılığı olunca, tüketicinin de satın alma süreçleri ve karar verme süreçleri de dikkatle incelenmektedir (Kotler, 2003).

Müşteri kaybı kavramını basit bir şekilde “Rekabet nedeni ile müşterinin firmayı tercih etmeyi bırakması” olarak tanımlayabiliriz (Nettleton, 2014). Bu durumdan da anlaşılacağı üzere eldeki müşteriye tutmak zorlaşmaktadır. Müşteri kaybının doğru belirlenmesi firmadan vazgeçip, vazgeçmemesi önem arz etmektedir. Bu yüzden müşteri bırakabilir veya bırakmaz şeklinde sınıflandırılmalıdır (Tsiptsis ve Chorionopoulos, 2009). Bu çalışmanın amacı müşteri profillerine bakarak firmayı hangi tür müşterilerin bıraktıklarını tahmin etmek ve buna karşı alınabilecek önlemleri ortaya çıkarmaktır.

Müşteri kaybı, genellikle hizmet işletmelerini ilgilendirmekte olup, başta telekomünikasyon ve bankacılık olmak üzere birçok müşteriye hizmet eden firmalar için müşterilerin tercihlerinin değişmesi veya firmayı tercihten vazgeçişinin artmasıyla gündem haline gelmiştir (Kim, Jun ve Lee, 2014). Firmalar yeni müşteri kazanabilmek için çeşitli yollara başvurmaktadır. Dolayısıyla maliyetleri de artırmaktadır. Bu yüzden yeni müşteri kazanmak için kampanyalar düzenleyip yeni müşteriler kazanmak yerine, elindeki müşteriye cezbedici ve bırakmaması için yeni yollara başvurmaktadırlar. En büyük avantaj ise müşteriye kaybetmeden önce belirlemek ve onun üzerine önlemler almaktır. İyi belirlenmiş bir modelleme ile hangi müşterinin bırakabileceği hangi müşterinin sadık olduğunu ortaya çıkarılabilir.

Gelişen teknoloji ile birlikte işletmelerin mevcut müşterilerinin durumunu analiz edebilmek amacıyla birçok çalışma yapılmaktadır. Örneğin; Tosun (2006), bir bankanın müşteri verileri üzerinde çalışmasını yaparak kredi kartı kullanmayı bırak müşterileri tespit etmeye çalışmaktadır. Çalışmasında karar ağacı yöntemlerini kullanmaktadır.

Akbulut (2006), bir kozmetik firmasının müşteri verilerini analiz etmektedir. Akbulut çalışmasında, WEKA kullanarak kümeleme ve sınıflandırma

yapmaktadır. Kullandığı veriler üzerinde en etkili sonucu J.48 karar ağacı ile elde ettiğini anlatmaktadır. Yaptığı analiz sonucunda müşterileri aldıkları ürünlere göre belirli kategorilere ayırmaktadır ve işletme için önemli olabilecek önerilerde bulunmaktadır.

Asilkan (2008), yapmış olduğu çalışmasında yapay sinir ağları kullanarak ikinci el otomobil fiyatlarının gelecek dönemlerdeki durumunu tahmin etmeye çalışmaktadır. Sonuçların karşılaştırılmasında zaman serisi analiz yöntemleri kullanılmıştır. Yapılan çalışmada, yapay sinir ağlarının kullanılmasının nedeni, eldeki az miktardaki veriden anlamlı sonuçların çıkarılması olarak tanımlanmaktadır.

Koçtürk (2010), bireysel emeklilik müşterilerinin kayıp analizini yapmaktadır. Çalışmada SAS Enterprise Miner kullanarak müşteri bağlılık davranışı modellenmiştir. Entropy tabanlı 21 yapraklı karar ağacı ile analiz yapılmıştır. En kaliteli değer sağlandığı 4 dallı ve 46 yapraklı hiyerarşinin sağladığı görülmüştür.

Arifoğlu (2011), bir GSM firmasının müşterilerine ait verileri analiz etmektedir. Çalışmada aynı veri seti ile Navie Bayes, Support Vector Machine, Probabilistic Neural Network ve C-means Algoritmaları karşılaştırılmıştır. Kullanılan veri seti için Adaptif Ağ Tabanlı Bulanık Çıkarım Sistemi (ANFIS) ile birlikte kullanılan kurallarda C-means Algoritması'nın en iyi sonucu verdiği görülmüştür.

Ercan (2015), çalışmasında tek bir veri seti kullanarak Bayesian Network, Logistic Regression, SMO ve Simple CART Algoritmaları'ndan elde ettiği sonuçları karşılaştırmakta ve yeni bir tahmin modeli oluşturmaktadır. Oluşturduğu yeni modelin doğruluğunu arttırmak için Ensemble Yöntemleri uygulanmaktadır. Modelden elde edilen doğruluk oranının %68,2 olduğu açıklanmaktadır.

Çimenli (2015), lojistik sektöründe faaliyet gösteren bir işletmeye ait müşteri verileri ile kayıp analizi yapmaktadır. Çalışmada yapay sinir ağları ve karar ağaçları karşılaştırılmıştır. Çalışma sonucunda karar ağaçlarından %81 doğru sonuç alınıyorken, yapay sinir ağlarından %97 doğru sonuç alındığı vurgulanmaktadır.

2. SINIFLANDIRMA YÖNTEMLERİNE GENEL BAKIŞ

Bu bölümde yukarıda bahsi geçen ve uygulamada kullanılan sınıflandırma yöntemleri hakkında bilgi verilmektedir.

Logistic Regression

Lojistik Regresyon, bağımlı ve bağımsız değişkenler arasındaki ilişkinin şeklini belirleyerek tahmin etmek ve sınıflandırmak amacıyla kullanılan bir algoritmadır (Teng, Lin ve Ho, 2007). Doğrusal Regresyon Algoritması'nın uygulanması için bazı koşulların sağlanması gerekmektedir. Bu koşullardan bir tanesi bağımlı değişkenin sürekli olmasıdır. Ancak, uygulamalarda bağımlı değişken genellikle kategorik ve iki düzeylidir. Bağımlı değişkenin iki düzeyli olduğu durumlarda Lojistik Regresyon Analizi kullanılmaktadır (Gujarati, 1999).

Naive Bayes

Naive Bayes (Denklem 1), Bayes Teoremi baz alınarak oluşturulan kolay uygulanabilirlik ve anlaşılabilirlik yönünden avantajlı olan basit makine öğrenme algoritmalarındandır. Bu algoritma niteliğin hangi sınıflandırma bölümüne ait olduğunu tahmin etmektedir.

Sınıflandırma

$$(s_1, s_2, \dots, s_n) = \text{argmax}_k p(K=k) \prod_{i=1}^n p(S_i = s_i / K=k) \quad (1)$$

1 numaralı denklem Naive Bayes Formülü'nden anlaşılacağı üzere S_1 ile S_n arasındaki sınıflardan seçim yaparken bu koşulları yerine getiren k çarpımından farkı yoktur. Her koşulun bir sınıf değeri vardır ve hangi koşulun hangi sınıftan olduğunu bulması için olasılık değeri çarpılır.

Decision Tree (CART)

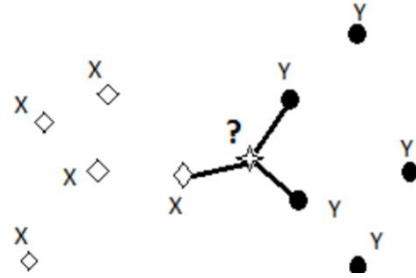
Sınıflandırma modellerinden olan karar ağaçları kolay kurulumu, etkili sonuçları ve oldukça fazla olan literatür çalışmasından dolayı en çok kullanılan yöntemdir. Karar ağaçları basit kararlar verebilme ve çok büyük verileri çok küçük sınıflara bölmek için kullanılır. Sıklıkla kullanılan karar ağacı modellerinden biri de CART Algoritması'dır. CART, kategorik sürekli değişkenli kullanarak sınıflama problemlerinin çözümünde karar ağaçlarını kullanan istatistiksel bir metottur. Bağımlı değişkenler kategorikse sınıflandırma ağaçları, eğer sürekliyse lojistik regresyon ağaçları olarak tanımlanır (Deconinck ve Hancock, 2005). CART, kümeyi ikiye ayırarak daha homojen ikili karar ağacı oluşturmak isteyen yapıya sahiptir. Bu yüzden CART veriyi iki düğüme ayırıp bağımsız değişkenleri kullanarak alt gruplara ayırmayı hedefler.

K-Nearest Neighborhood (K-NN)

Sınıflandırma işleminde temel amaçlardan biri en basit haliyle nesnelerin özelliklerine bakılarak nesnelerin hangi sınıfa ait olduğu belirlenmesidir. Çeşitli sınıflandırma algoritmaları vardır. Bunların içinde en çok ve yaygın kullanılan makine öğrenim algoritması (Denklem 2 ve Şekil 1) KNN Algoritmasıdır. Bir özellik seçilerek ona en yakın özellikle arasındaki yakınlığa göre sınıflandırma yapar ve nesneler arasındaki mesafelerin hesaplanmasında kullanılır.

$$d_{(i,j)} = \sqrt{\sum_{k=1}^p (X_{ik} - X_{jk})^2} \quad (2)$$

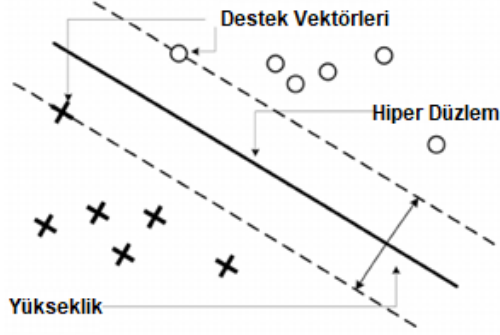
Çalışma prensibine bakıldığı zaman yeni bir nesne sınıflandırılmak için k değerine bakılır. Burada herhangi bir eşitlik olmaması için k değeri tek sayı olarak belirlenir. Yeni veri ile var olan veri hesaplanırken Manhattan uzaklığı veya Öklid gibi yöntemler kullanılır (Kılınç, Borandağ, Yücelar, Tunali, Şimşek ve Özçift, 2016).



Şekil 1. K-Nearest Neighborhood (K-En yakın komşu Algoritması)

Support Vector Machine (SVM)

Sınıflandırma sorunlarını çözmek için son yıllarda geliştirilen makine öğrenimi algoritmalarından biridir. Aynı zamanda örüntü tanımadaki popüler bir genelleme yöntemidir (Ayhan ve Erdoğan, 2014). Bu problemlerin çözümünde başarıyla uygulanmış ve yüksek başarımla elde edilmiş makine öğrenimi algoritmalarından biridir (Sengur, 2019). Veri madenciliği uygulamalarında sınıflandırma, doğrusal veya doğrusal olmayan bir işlev ile yerine getirilmektedir. Kernel İşlevi ile doğrusal olmayan dönüşümlerde, verileri doğrusal olarak ayırmaya olanak tanınmış olunur. Şekil 2'de Destek Vektör Makine Algoritması'nın genel yapısı verilmiştir (Karagülle, 2008).



Şekil 2. Destek Vektör Makine Algoritması'nın genel yapısı

1960'lı yıllara dayanan istatistiksel öğrenme algoritmalarından biri olan Destek Vektör Makinesi, Vapnik ve Chervoneskis tarafından tanımlanmış ve çok sayıda problemi başarıyla çözmüştür. Orijinal ifadesi Support Vektör Machine (SVM) olan Destek Vektör Makine Algoritması'nın temel mantığı, verileri doğrusal olarak ayırıştırabilen en iyi çözümün belirlenmesidir (Sezer, Erçil ve Keskinöz, 2005). Doğrusal olarak ayırıştırılmayan veriler dönüşüm tekniği ile farklı bir boyuta taşınarak çözülür. SVM'lerde kullanılacak örnek sayısı mühim değildir. SVM öğrenim sırasında görülmemiş verileri de sınıflayabilmektedir. Bunu genelleyme özelliği sayesinde yapabilmektedir. Bu özelliği diğer tekniklerden ayıran en önemlisi olarak görülmektedir. Son yıllarda yüz tanıma, veri madenciliği, biyoloji, gen analizi ve örüntü tanıma sınıflandırılmalarında SVM kullanılmaya başlanmıştır. Bir olayı tanımlamak için özellikler vektör ile çağrılarak doğru seçilmelidir. Bu sayede doğru sonuçlar elde edilebilmektedir (Kecman).

Linear Discriminant Analysis (LDA)

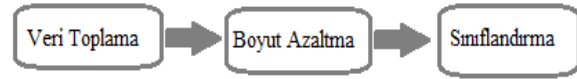
Linear Discriminant Analysis (LDA), verilerin sayısallaştırılmadığı modellerde kaplanan boyutu küçültmek, hesaplama verimliliğini artırarak algoritma başarı oranını yükseltmek için kullanılır (Martinez ve Kak, 2001). LDA, sınıflandırmayı en iyi duruma getiren özellik alt uzayını bulmaktır. LDA, bir veri kümesindeki boyut sayısını azaltmak ve sınıflandırmak için kullanılan doğrusal dönüşüm teknikleridir. LDA sonucu iki sınıf oluşur (Hua ve Yang, 2001).

LDA, bağımsız değişkenlerin aralarında nasıl bir etkileşim olduğunu ve bu etkileşimin bağımlı değişkeni nasıl etkilediğini ortaya çıkarmaya çalışan bir Varyans Analizi (ANOVA) ile yakından ilgilidir (Fisher, 1936). Kısaca LDA verilerin sınıflandırılmasından sonra bu

verilerin arasındaki farkı tespit etmeye çalışır (Perriere ve Thioulouse, 2003).

3. ANALİZ METODU

Kullanılan sınıflandırma algoritmaları: Logistic Regression, Naive Bayes, Decision Tree, K-NN, SVM ve LDA'dır. Verilerin analizinde Python ortamının seçilmesinin sebebi algoritma üzerinde değişiklik yapılabilmesi ve yapılan literatür çalışmaları etkili olmuştur. Yapılan çalışmada sınıflandırma modelleri ile müşteri kayıp tahmini yapılmak istenmiştir. Çalışma sonunda elde edilen değerler müşterinin kayıp edilip edilmediğini göstermektedir. İzlenen yöntemler Şekil 3'teki gibi özetlenebilir.



Şekil 3. Yöntem

3.1. Veri Toplama

Analizlerde kullanılacak veriler, şirketlerin belirli bilgilerini ya da çalışma verileri yayımlayan kaggle.com'dan temin edilmiştir. Veri setinin %80'ni eğitim, %20'si test için kullanılmıştır. Tablo 1'de özellikleri verilen veri seti bir bankaya aittir ve müşteri bilgilerinin içermektedir. Tablo 2'de bulunan veri seti ise bir operatör firmasına ait verilerden oluşmaktadır. Tabloda veri setinde bulunan bütün özellikler yer almaktadır. Tablo 1'de görünen ilk seti için boyut azaltma işlemi yapılarak önemli olduğu bilinen özelliklerden faydalanılmıştır.

Tablo 1. Banka veri setindeki özellikler

Özellik	En Küçük Değer	En Büyük Değer
Satır Sayısı	1	10000
Müşteri Numarası	15528523	15745083
Soyadı	-	-
Kredi Skoru	490	800
Yaşadığı Ülke	-	-
Cinsiyet	-	-
Yaş	24	58
Firma ile Çalıştığı Süre (Yıl)	0	13
Hesap Tutarı	10134,88	99183,9
Kullandığı Ürün Sayısı (Kart Sayısı vb.)	1	9

Kredi Kartı Olması Durumu	0	1
Firma ile Aktif Çalışma Durumu	0	1
Tahmini Maaş	2924,56	92865,41

Tablo 2. Telefon operatörü veri setindeki özellikler

Özellik
Durum
Firma ile Çalışılan Süre
Alan Kodu
Telefon Numarası
Uluslararası Plan
Sesli Posta Planı
Vmail Mesajları
Toplam Gün
Toplam Yapılan Günlük Arama
Toplam Yapılan Günlük Ücret
Toplam Yapılan Ev Telefonu Dakikası
Toplam Yapılan Ev Telefonu Araması Sayısı
Toplam Yapılan Ev Telefonu Araması Ücreti
Toplam Yapılan Gece Görüşme Dakikası
Toplam Yapılan Gece Görüşme Ücreti
Toplam Yapılan İnternet Dakikası
Toplam Yapılan Arama
Toplam Yapılan Ücret
Toplam Müşteri Hizmetleri Çağrı Sayısı

3.2. Boyut Azaltma

Tablo 1’de yer alan veri seti için boyut azaltma çalışmasından sonra kullanılan özelliklerin değişimi Tablo 3’te bulunmaktadır.

Tablo 3. Banka veri setinde kullanılan özellikler

Özellik	En Küçük Değer	En Büyük Değer
Kredi Skoru	490	800
Yaşadığı Ülke	-	-
Cinsiyet	-	-
Yaş	24	58
Firma ile Çalıştığı Süre (Yıl)	0	13
Hesap Tutarı	10134,88	99183,9
Kullandığı Ürün Sayısı (Kart Sayısı vb.)	1	9
Kredi Kartı Olması Durumu	0	1
Firma ile Çalıştığı Süre (Yıl)	0	13
Tahmini Maaş	2924,56	92865,41

3.3. Sınıflandırma

Sınıflandırma işlemlerinden sonra her müşteri özelinde bir sonuç beklenmektedir. Kullanılan sınıflandırma Tablo 4’te yer almaktadır.

Tablo 4. Sınıflar

Sonuç	Sınıf (Class)
0	Kaybedilen Müşteri
1	Kazanılan Müşteri

4. SONUÇ

Belirlenen tüm sınıflandırma algoritmaları ile iki veri seti üzerinde de çalıştırılmıştır. Bankaya ait olan ve üzerinde boyut azaltma işlemi yapılan Tablo 3’teki veri setinden, Tablo 5’teki sonuçlar elde edilmiştir. Elde edilen sonuçlara göre ilk veri seti için en iyi sonucu %81,00’lik doğruluk oranı ile LDA ve %76,40’lık bir değer ile en az doğruluk oranı olan algoritma K-NN’dir.

Tablo 5. Banka veri setinden elde edilen sonuçlar

Algoritma	Accuracy Score
LDA	%81,00
Decision Tree	%80,30
Logistic Regression	%79,75
SVM	%79,75
Naive Bayes	%78,45
K-NN	%76,40

Tablo 2’de yer alan bir operatör firmasına ait veri seti ile elde edilen sonuçlar Tablo 6’da gösterilmektedir. Tablo 6’da görüldüğü gibi bu veri seti için en iyi sonucu %80,4’lik oran ile Decision Tree algoritması ile elde edebilmekteyiz. En düşük doğruluk oranı ise %72,16 ile LDA ve Logistic Regression algoritmalarından elde edilmiştir.

Tablo 6. Operatör firmanın veri setinden elde edilen sonuçlar

Algoritma	Accuracy Score
Decision Tree	%80,41
Naive Bayes	%76,80
K-NN	%74,74
SVM	%74,23
LDA	%72,16
Logistic Regression	%72,16

Tablo 5'te ve Tablo 6'da görünen performans sonuçlarına göre uygulamalarda kullanılacak yöntemler ve veri setleri arasında doğru orantı olduğunu söylemek mümkündür.

5. TARTIŞMA

Sınıflandırma yaparken kullanılan algoritmaların seçimi, kullanılacak veri tipiyle uyumlu olmalıdır. Böylece daha doğru sonuçlar elde edebilmemiz mümkündür. İki veri seti de sınıflandırma ile uyumlu olmasına rağmen bazı algoritmalar iki veri seti için de beklenenden daha farklı sonuçlar vermiştir. Bunun sebebi iki veri setindeki özelliklerin sayıca fazlalığından olabilir. Çalışmada, 1. veri setinde (Tablo 1 Banka veri seti) 10 özellik var iken 2. veri setinde (Tablo 2 Telefon operatörü veri seti) 19 özellik vardır. İki veri seti için de özellik sayılarının birbirinden farklı olması, sınıflandırma algoritmalarında yaklaşık olarak %10'luk farklı bir oran vermiştir. Algoritmalarından elde edilen sonuçlar aşağıdaki gibi değerlendirilmiştir.

- LDA Algoritması'nda daha fazla özelliğe sahip olan 2. veri setinin başarı oranı daha yüksek iken daha az özelliğe sahip olan 1. veri setinde en az doğruluk oranı sağlayan 2. algoritma olmuştur.
- K-NN Algoritması ise daha az özelliğe sahip olan 1. veri seti için en az doğruluk oranına sahip algoritma olmuştur. 2. veri seti için ise, Tablo 6'da görüleceği üzere ortalama bir doğruluk oranı elde etmiştir.
- Decision Tree Algoritması, iki veri seti için de yaklaşık olarak %80 doğruluk oranı vermiştir.
- Naive Bayes Algoritması, 1. veri setin en az doğruluk oranına sahip algoritmalarından biri olurken, 2. veri seti için diğer algoritmalara göre daha yüksek oranda doğruluk elde edilmiştir.
- SVM, iki veri setin de karşılaştırma tablolarında ortalama sıralamada yer almaktadır. Bunun sonucu olarak; seçilen veri setlerinin SVM için daha uyumlu olabileceği düşünülmektedir.
- Logistic Regresyon, 1. veri seti için daha iyi sonuç verirken, 2. veri seti için en az doğruluk oranı veren algoritma olmuştur.

Verinin analizinin güvenilirliği, özelliklerinin iyi belirlenmesi ve sınıflandırmanın tam yapılması algoritmanın başarısını olumlu yönde etkileyeceği düşünülmektedir. Her iki veri seti için de elde edilen sonuçların %70'ten fazla olması kullanılan algoritmaların veri setleri için uygun olduğunu göstermektedir. Bu nedenle Tablo 5'te ve Tablo 6'da gösterilen sonuçlarda algoritmaların doğruluk oranları farklı olsa da, yapılan iki çalışma da genel olarak başarılı kabul edilebilir.

KAYNAKLAR

Akbulut S. (2006). *Veri madenciliği teknikleri ile bir kozmetik markanın ayrılan müşteri analizi ve müşteri segmentasyonu*, (Yüksek lisans tezi). Gazi Üniversitesi Fen Bilimleri Enstitüsü, Ankara.

Arifoğlu E. (2011). *Churn management by using Fuzzy C-Means*, (Yüksek lisans tezi). Bahçeşehir Üniversitesi Fen Bilimleri Enstitüsü, İstanbul.

Asilkan Ö. (2008). *Veri madenciliği kullanılarak ikinci el otomobil pazarında fiyat tahmini*, (Doktora tezi). Akdeniz Üniversitesi Sosyal Bilimler Enstitüsü, Antalya.

Ayhan, S. ve Erdoğan, Ş. (2014). Destek vektör makineleriyle sınıflandırma problemlerinin çözümü için çekirdek fonksiyonu seçimi. *İibf Dergisi*, 9(1), 175-198.

Çimenli S. (2015). *Churn analysis and Prediction with decision tree and artificial neural network*, (Yüksek lisans tezi). Kadir Has Üniversitesi Fen Bilimleri Enstitüsü, İstanbul.

Deconinck, E., Hancock, T., Coomans, D., Massart, D.L. and Heyden, Y.V. (2005). Classification of drugs in absorption classes using the classification and regression trees (CART) methodology. *Journal of Pharmaceutical and Biomedical Analysis*, 39, 91-103.

Ercan P. (2015). *Detection of churners in internet games using crm approach: a case study on pishti plus*, (Yüksek lisans tezi). Orta Doğu Teknik Üniversitesi Fen Bilimleri Enstitüsü, Ankara.

Fisher, R. A. (1936). The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, 7(2), 179-188.

Gujarati, D.N. (1999). *Temel ekonometri*, (Şenesen, Ü. ve Şenesen, G.G., Çev.), İstanbul: Literatür Yayıncılık (Orijinal çalışma basım tarihi 1978).

Hua, Y. and Yang, J. (2001). A direct LDA algorithm for high-dimensional data-with application to face recognition, *Pattern Recognition*.

Karagülle, F. (2008). *Destek vektör makinelerin kullanarak yüz bulma*. (Yüksek lisans tezi).

Kecman, V. (2001). *Learning and Support Vector Machines Neural Networks And Fuzzy Logic Models*.

Kılınç, D., Borandağ, E., Yücelar, E., Tunalı, V., Şimşek, M. ve Özçiftçi, A. (2016). KNN algoritması ve R dili ile metin madenciliği kullanılarak bilimsel makale tasnifi. *Marmara Fen Bilimleri Dergisi*, (3), 89-94.

Kim, N., Jung, K. H., Kim, Y. S., and Lee, J. (2012). Uniformly subsampled ensemble (USE) for churn management: *Theory and implementation*. *Expert Systems with Applications*, 39(15), 11839–11845.

Koçtürk Y. (2010). *Veri madenciliğinde bağlılık*, (Yüksek lisans tezi). İstanbul Teknik Üniversitesi Fen Bilimleri Enstitüsü, İstanbul.

Kotler, P. (2003). *Marketing Insights from A to Z*. New Jersey.

M. Accardi-Petersen. (2011). Agile marketing. New York: Apress. Commercial data mining: processing, *Analysis and Modeling for Predictive Analytics Projects*.

Martinez, A. M. and Kak, A. C. (2001). PCA versus LDA(PDF). *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(2), 228-233.

Perriere, G. and Thioulouse, J. (2003). Use of correspondence discriminant analysis to predict the subcellular location of bacterial proteins. *Computer Methods and Programs in Biomedicine*, 70, 99–105.

Sengur, A. (2019). Ulticlass least-squares support vector machines for analog modulation classification, *expert systems with applications*, 36(3), 6681-6685.

Sezer, O., Erçil, A. ve Keskinöz, M. (2005). Destek vektör makinesi kullanarak bağımsız bileşen tabanlı 3B nesne tanıma, *Sabancı University Computer Vision and Pattern Analysis Laboratory*.

Teng, J., Lin, K. and Ho, B. (2007). Application of classical tree and logistic regression for the management and health intervention plans in a community-based study, *Journal of Evaluation in Clinical Practice*, 13, 741-748.

Tosun T. (2006). *Veri madenciliği teknikleriyle kredi kartlarında müşteri kaybetme analizi* (Yüksek lisans tezi). İstanbul Teknik Üniversitesi Fen Bilimleri Enstitüsü, İstanbul.

Tsiptsis, K., and Chorianopoulos, A. (2009). Data mining techniques in CRM: *Inside Customer Segmentation*. John Wiley & Sons, Ltd.