

Karıncı Kolonisi Optimizasyonu Sınıflandırma Algoritması Yöntemi İle Telefon Bankacılığında Doğrudan Pazarlama Kampanyası Üzerine Bir Sınıflandırma Analizi

A classification analysis on the direct marketing campaigns in telephone banking with ant colony optimization classification algorithm method

Eyüp AKÇETİN¹, e.akcetin@gmail.com

Ufuk ÇELİK², ucelik001@gmail.com

Geliş Tarihi/Received: 10.02.2015; **Kabul Tarihi/Accepted:** 27.04.2015

doi: 10.5505/iuyd.2015.20592

Telefon bankacılığında doğrudan pazarlama son zamanlarda önem arz eden bir konu olmuştur. Bu tür kampanyaların başarısı müşterilerin etkili bir şekilde katılımları ile doğru orantılıdır. Özellikle bankalar tarafından çok yürütülen bu tür kampanyaların başarısı veri madenciliği yöntemleri ile artırılabilir. Bu çalışmanın temel amacı, veri madenciliği yöntemlerinden sezgisel bir algoritma olan karıncı kolonisi optimizasyonu ile sınıflandırma işleminin telefon bankacılığı verileri için geliştirilmesidir. Analiz için Portekiz'deki bir bankanın gerçek verileri kullanılmıştır. Karıncı kolonisi algoritması kullanılarak müşterilerin kredi alabilme ihtimalleri tahmin edilmiş ve sonuçlar veri madenciliğinde yaygın olarak kullanılan diğer yöntemler ile karşılaştırılmıştır. Başarı oranı sınıflama doğruluğu, hassasiyet ve kesinlik değerleri ile hesaplanmıştır.

Anahtar Kelimeler: Veri madenciliği, Karıncı Kolonisi Optimizasyonu Sınıflandırma Algoritması, Telefon Bankacılığı, Müşteri analizi.

Jel Kodları: C80, C61, D81, Y10, M31.

Direct marketing in telephone banking has been an important topic recently. The success of such campaigns is directly proportional to the effective customer participation. The success rate of such campaigns that especially conducted by banks may be enhanced by data mining methods. The main purpose of this study is to implement the classification process via ant colony optimization which is a heuristic algorithm in data mining for telephone banking dataset. The real world data from a bank in Portuguese for analysis. The possibilities of the customer credit subscription were estimated by using ant colony optimization algorithm and the results were compared to other commonly used methods in data mining. The rate of the success were calculated by classification accuracy, sensitivity and specificity.

Keywords: Data mining, Classification of Ant Colony Optimization Classification Algorithm, Telephone Banking, Customer analysis.

Jel Codes: C80, C61, D81, Y10, M31.

¹ Yard. Doç. Dr., Balıkesir Üniversitesi, Bandırma Denizcilik Fakültesi (Yazışılan yazar)

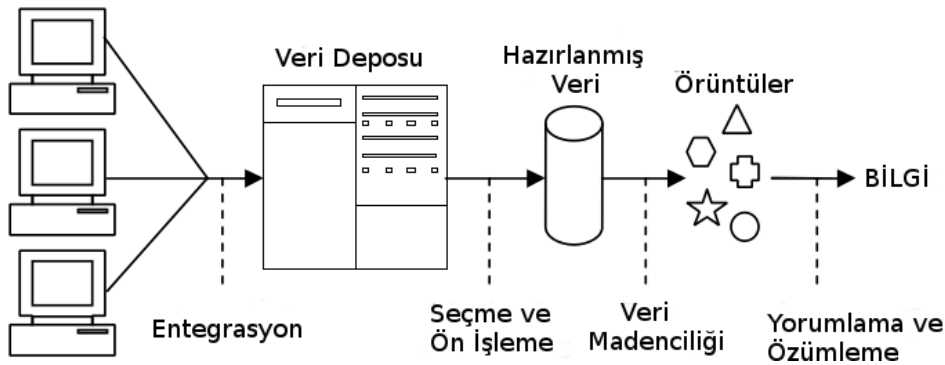
² Öğr. Gör., Balıkesir Üniversitesi, Gönen MYO

1. GİRİŞ

Günümüzde çok büyük hacimlerdeki veriler bankalar ve diğer finansal kuruluşlar tarafından depolanmaktadır. Bu veriler içindeki değerli bilgileri ortaya çıkarmak işletme karlılığı açısından çok önemlidir. Bu nedenle bu veriler üzerinde ciddi analizlerin yapılması gereklidir. Bankacılık sektöründe veri madenciliği; pazarlama, ticari analiz, portföy yönetimi, risk yönetimi, müşteri ilişkileri yönetimi, müşteri kazanımı ve elde tutulması, müşteri profili analizi ve müşteri hizmetleri alanında iş performansını artırmak amacı ile kullanılmaktadır. Veri madenciliği yöntemlerinin bankacılık sektöründe kullanılmasının iki temel sebebi şudur. Veri ambarlarında biriken yoğun veri kümesi içinden az ama önemli bilgiye hızlı, etkin bir şekilde ulaşmak ve işletme için faydalı bilgileri ortaya çıkarıp yorumlamaktır. (Gürsoy, 2009: 153-166).

Veri madenciliğinin verimli ve etkin analizlerde kullanılması için öncelikle iyi tasarlanmış veri ambarlarına ihtiyaç vardır. Bu veri ambarları her türlü veriyi sistematik olarak içerecek yapıda olmalıdır. Veri ambarı, konu odaklı, bütünleştirilmiş, zaman içinde değişebilen, süreklilik arz eden veri toplama işlevi ile karar verme süreçlerine destek sağlayan sistem ve/veya sistemler bütünüdür. Veri madenciliği ise veri ambarından değerli bilginin çeşitli yöntemlerle çıkarılması sürecidir (Özkan, 2013: 37-50). Günümüzde veri ambarlarından verilerin analiz edilerek önemli kurumsal bilgilere dönüştürülmesinde yeni teknolojiler çok etkin hale gelmiştir (Silahtaroglu, 2013: 17-18). Veri madenciliği ispatlanması veya çürütülmesi gereken bir hipotez ile başlamaz. Veri madenciliği, geleneksel bilgi doğrulamadan ziyade bilgi keşfetme amaçlı araştırma sürecidir. Veri madenciliği beklenmeden modelleri ve fırsatları ortaya çıkararak kendi hipotezini oluşturur. Veri ambarı, veriyi depolayıp, biçimleyip ve organize ederken veri madenciliği ise işletme açısından yararlı bilgileri ortaya çıkarıp trendleri öngörmeyi hedefler (Köktürk & Dirsehan, 2012: 3). Veri madenciliğinde bilgi edinme süreci (Bramer, 2013) Şekil 1 ile gösterilmiştir.

Veri Kaynakları



Şekil 1. Veri madenciliğinde bilgi edinme süreci

Veri madenciliğinde bilgi edinme amaçlı olarak sınıflandırma çok sık uygulanan bir yöntemdir. Sınıflama için kural tabanlı algoritmalar, karar ağaçları, bayes yöntemi, geri yayılım algoritmaları, karar destek sistemleri ve bulanık mantık gibi yöntemler kullanılır (Han & Kamber: 2006). Özellikle İnternet bankacılığının yaygınlaşması ile bankalarda sınıflandırma amaçlı veri madenciliği kullanımı artmıştır.

Bu çalışmada ise aslında optimizasyon işlemlerinde kullanılan sezgisel algoritmalarından karınca kolonisi optimizasyonu sınıflandırma algoritması (KKOSA) bahsedilecektir. İkinci bölümde yazın taraması ile bankacılık sektöründe veri madenciliği ile yapılan çalışmalar anlatılarak, üçüncü bölümde KKOSA kullanım amacıyla, dördüncü bölümde algoritma metodolojisi ve materyalden bahsedilecektir. Beşinci bölümde bulgular ve çıkarımlardan bahsedilerek altıncı bölümde sonuçlar analiz edilecektir.

2. LİTERATÜR TARAMASI

Veri madenciliği yöntemleri ilk önce marketlerden yapılan alışverişlerin analizinde kullanılmıştır. Marketler bu yöntemleri kullanarak yapılan alışverişlerde birlikte alınan ürünleri inceleyerek birlikte alınan ürünleri market raflarında birbirlerine yakın noktalara konumlandırmışlardır. Yapılan promosyonlarda bu ürünler birlikte kullanılarak müşterilerin dikkati çekilmeye ve tüketim faaliyetleri artırılmaya çalışılmıştır. Ayrıca bu analizler optimum stok ve envanter yönetimi içinde gerekli bilgiyi de sağlamışlardır. ÖrneğinWalMart'ın yaklaşık 100 milyon müşterisi haftada 3600'den fazla markete alışveriş için gelmektedir (Han vd., 2006: 675). WalMart veri madenciliği yöntemlerini müşteri ve tedarikçi ilişkilerini geliştirmek için kullanmaktadır. Günlük yaklaşık olarak 2900 marketten 7,5 terabayt veriyi analiz ederek ürünleri raf raf ve tek tek izlemektedir. Böylece market ve tedarikçi ilişkilerini çok yakından takip etmektedir. 1995 yılında WalMart'ın bilgisayarlarında 1 milyondan fazla veri sorgulaması yapılmıştır. WalMart 3500'den fazla tedarikçisine veri ambarlarına ulaşarak veri analizi yapmasına izin vermektedir. Tedarikçiler bu analizlerle, müşterilerin alışveriş davranışlarını inceleyerek her markete özgü envanter planı yapmakta ve yeni pazarlama stratejileri yakalamaktadırlar (Wijaya, 2013).WalMart veri madenciliği yöntemlerini kapsamlı şekilde kullanarak satışlarını %20 oranında artırmayı başarmıştır. Bazı ürünlerin birlikte alınma olasılığının tahmin edilmesi kolayken bazı ürünlerin birlikte alınma olasılığının tahmini zordur. Örneğin; dış macunu ile dış fırçası, çay ile şeker birlikte alınan ürünlerdir. Bu tahmini yapmak kolay olmasına rağmen bazı spesifik tahminler oldukça zordur. Örneğin; WalMart'ta Cuma akşamları bebek bezi ile biranın bir arada satıldığını tahmin etmek ve bu iki ürünü market içinde birbirine yakın noktalara stoklamak ilk bakışta mantıksız görünebilir. Fakat veri madenciliğinde amaç büyük verilerin analizi yapılarak bu tür tahmini zor örnekleri çıkarıp işletme verimliliğini artırmaktır. Günümüzde veri, tıpkı petrol gibi çeşitli işlemlerden geçmektedir Veri madenciliğinde önemli olan veri tabanındaki verinin, çeşitli analizler sonucunda kurumsal açıdan stratejik ve değerli bilgilere dönüştürülmesidir. Verinin bilgiye dönüşümü, işletmelere, rakipleri karşısındarekabetçi üstünlük sağlamaktadır (Domingo, 2015). Bu nedenle veri madenciliği yöntemleri, stratejik bilgilere erişmek ve rekabetçi üstünlük elde edebilmek amacı ile sağlıktan bankacılığa birçok sektör tarafından etkin biçimde kullanılmaktadır.

Bankacılık hizmetlerinin pazarlanmasında veri madenciliği, pazarlama tekniklerinin geliştirilmesinde ve hedef müşterilerin tespitinde kullanılmaktadır. Kullanılan yöntemlerle müşteri bölümlendirme yapılarak daha önce yapılan kampanyalara iştirak eden müşteriler belirlenir. Böylece hangi kampanya hangi tür müşterilerin iştirak edeceği tahmin edilir. Çapraz satış için karlı müşterilerin profili çıkarılır ve onlara uygun ürünler geliştirilebilir. Ayrıca kurumsal yıpranma analizi yapılarak müşteri davranışlarındaki sapmalar tespit edilebilir (Bach vd., 2013: 32-41).

Bankacılık sektöründe, risk yönetimi çok önemli bir konudur (Liébana-Cabanillas vd., 2013: 5439-5447). Veri madenciliği yöntemlerinin detaylı müşteri analizleri sayesinde, müşterilerin veya tüketicilerin bağlı olduğu finans kurumuna vereceği zararlar en aza indirgenebilir ve yüksek riskli müşteriler elimine edilebilir (Gray & Debrecey, 2014: 357-380).

Avrupa Birliğinde, 2010 yılında kredi kartı ve diğer kart dolandırıcılığının toplam parasal seviyesi 1,26 milyar Euro'ya ulaşmıştır (European Central Bank, 2012: 4). Kredi kartının kaybolması, üçüncü şahıslar tarafından kullanılması veya çalınması durumunda oluşacak zararlar, bankalar tarafından karşılanmaktadır. Bu tür sebeplerle oluşabilecek zararları azaltmak için risk azaltıcı ve sahtekârlık tespit edici sistemler bankalar tarafından oluşturulmaktadır (Argüden & Erşahin, 2008). Finansal dolandırıcılık ve sahte bankacılık işlemlerinin tespitinde veri madenciliği yöntemleri son yıllarda finansal kuruluşlar tarafından etkin biçimde kullanılmaktadır (Ravisankara vd., 2011: 491-500).

Finansal kurumlar ve diğer işletmeler veri madenciliği yöntemlerini, daha iyi karar verebilmek ve iş kararlarına yardımcı olmak amacıyla, verilerin desenleri ile ilişkilerini keşfetmek için kullanmaktadır. Veri madenciliği satış eğilimlerinin belirlenmesine, doğru pazarlama kampanyalarının geliştirilmesine ve müşteri sadakatinin kesinleştirilmesinde işletmelere yardımcı olmaktadır (Sumathi & Sivanandam, 2006: 16-17).

Bankalar, aynı ürünleri ve hizmetleri talep eden müşterilerinin ortak özelliklerini veri madenciliği yöntemlerini kullanarak tanımlayabilir. Marketsepeti analizi yapılarak hangi ürünlerin birlikte ve ne zaman satın alındığına yönelik detaylı bilgilere ulaşılır. Böylece müşterilere uygun ürün ve servis oluşturulurken, ürün ve servislere uygun müşteriler bulunabilir (Brito vd., 2015: 1-8).

Rakip bankaları tercih edecek mevcut müşterilerin önceden tahmin edilesi ve gerekli önlemlerin alınarak müşteri kayıplarının en aza indirgenmesinde veri madenciliği yöntemleri sık tercih edilmektedir (Keramati vd., 2014: 994-1012). Özellikle banka açısından değerli müşterilerin devamının sağlanmasında bu tür yöntemler önemlidir. (Ali & Arıtürk, 2014: 7789-7903).

Doğrudan pazarlama için tüm müşterilere tek tek çeşitli iletişim kanalları ile ulaşılması, işletmelere büyük mali külfet getirecektir. Bu nedenle müşteri profillerinin veri madenciliği yöntemleri ile analiz edilerek en doğru ürünün en doğru müşteriye pazarlaması için telefonla fazladan aranacak müşteri sayısının azaltılması karlılığı, önemli oranda artıracaktır (Javaheri & Teimourpour, 2014: 153-180) (Liao & Hsieh, 2011: 6059-6069). Yapılan bu analizler işletmelere stratejik ve taktiksel güç kazandırarak rekabetçi üstünlük sağlayacaktır (Crone & Stahlbock, 2006: 781-800).

Etkileşimli pazarlamada veri madenciliği yönteminin kullanılması özellikle İnternet bankacılığını kullanan müşterilerin ne tür işlemler yaptığı ve hangi ürünler ile servisleri kullandığı analiz edilebilir. Sonrasında müşteri ilişkilerin geliştirilerek müşterilere özel hizmet sunulabilir (Shaw vd., 2001: 127-137).

Ayrıca veri madenciliği trend analizlerinin yapılmasında kullanılmaktadır. Böylece önceki zaman dilimlerinde yapılan satışlar incelenerek gelecek zaman dilimlerine yönelik tahminde bulunulabilir (Oz, 2006: 355).

Bankacılık sektörünün son dönemlerde üzerine düştüğü önemli bir konuda telefon bankacılığıdır. Telefon bankacılığı ile müşterilerin kredi verilebilirlik durumları veri madenciliği yöntemleri ile tespit edilmektedir. Bu çalışmada kullanılan veri seti ile yapılan diğer çalışmalarda(Hany & Alaa, 2013) (Nachev, 2014) yapay sinir ağları ile sınıflandırma denenmiş ve %90 doğruluk oranı bulunmuştur. Karar ağaçları ile yapılan diğer çalışmada ise(Kumari & Shrivastava, 2015) %88.67 doğruluk oranı tespit ederken, çok hızlı karar kuralları algoritması (Kosina & Gama, 2015) ile yapılan sınıflandırma çalışmalarında %87 doğruluk oranı bulunmuştur. Ayrıca sezgisel bir yöntem olan arı kolonisi algoritması ile bu veri setini kullanarak kümeleme çalışması (Kuo vd., 2014) yapılmıştır. Aynı veri setinin kullanıldığı diğer bir çalışmada (Elsalamony, 2014) Naive Bayes, yapay sinir ağları, C5 karar ağacı modelleri performansları test edilirken diğer bir çalışmada(Popelka vd., 2012) RandomForest, ADTree, BFTree, C4.5 gibi karar ağaçları ile yapay sinir ağları ve genetik algoritmalar kıyaslanmıştır. Diğer bir çalışmada(Govindarajan, 2014) ise Radyal Tabanlı Fonksiyon ile Destek Vektör Makinesi algoritmalarımelezleştirilmesi ile %88 doğruluk oranına ulaşılmıştır.

3. AMAÇ

Veri madenciliği diğer sektörler gibi bankacılık alanında da iyi karar verme için kullanılır. Eş zamanlı olarak kötü niyetli işlemlerin ve kişilerin tespit edilmesinde yardımcı olur. Doğru amaçların ve yeni müşterilerin bulunmasında yardımcı olur. Müşteri özelliklerinin detaylı analiz edilmesini sağlar. Alışveriş örnekleri ile müşterilerin zamanla daha iyi tanınması sağlanarak müşteriye uygun servis ve ürünler müşteriye sunulur ve müşteri memnun edilerek, müşteri sadakati artırılır.

Kısaca bankacılık sektöründe veri madenciliği aşağıda sıralanan sorulara yanıt vermeye çalışır.

- Müşterinin sadakati nasıl sağlanır?
- Sadık müşteriler ne tür ürünleri ve servisleri talep etmektedir?
- Müşteri gözünde işletmeyi yıpratıcı unsurlar nelerdir?
- Bir müşteri rakip bankalar ile çalışmaya başlamadan önce hangi işlemleri yapmaktadır?
- Çapraz satış için, ATM kullanan müşterilerin profili nedir?
- ATM kullanan müşteriler genelde hangi ürünleri ve/veya servisleri kullanmaktadır?
- Çapraz satış ve diğer ürünler için müşterilerin belirlenmesi için hangi müşteri grubu hangi bankanın ürünlerini birlikte kullanmaktadır?
- Kredi kartı dolandırıcılığında yapılan işlemler nelerdir?
- Yüksek riskli tüketiciler ve müşteriler kimlerdir? Belirgin özellikleri nelerdir? vb.

Telefon bankacılığı müşteri kazanımı ve bilgi dönüşü açısından önemli bir çalışma yöntemidir. Müşterilerin doğru bir şekilde analiz edilebilmesi için alınan bilgilerin veri madenciliği ile araştırılması kayda değer sonuçlar ortaya çıkaracaktır. Bu bağlamda veri madenciliği yöntemleri ile müşterilerin tespit edilmesi daha etkin kararlar alınmasını sağlayacaktır.

Bu çalışmada KKOSA yönteminin yaygın kullanılan diğer veri madenciliği yöntemleri ile kıyaslanması ve mobil ve sabit telefon bankacılığı verileri üzerinde test edilmesi

amaçlanmıştır. Böylece yukarıda sıralanan sorulara yanıt aranırken KKOSA yönteminin finans veya diğer sektörlerde verimli kullanılabilirliği test edilmiştir.

4. METODOLOJİ VE MATERYAL

Mobil ve sabit telefon bankacılığı kampanyası analizi için Kaliforniya Üniversitesi makine öğrenmesi veri setlerinden Bank Marketing Data Set (Moro vd., 2011) isimli veri setinde, toplamda 45211 adet ve hızlı bir algoritma analizi için seçilmiş olan %10'luk 4521 adet veri kullanılmıştır. Bu veriler Portekiz Bankacılık Kurumu'nun Mayıs 2008 ile Kasım 2010 arasında elde ettiği mobil ve sabit telefonlara dayalı doğrudan pazarlama kampanyaları ile ilgilidir. Bu verilerin sınıflandırılmasında hedef müşterinin vadeli mevduat uygunluğunun tahmin edilmesidir. Banka vadeli mevduat ürünü verilir verilemeyeceğinin anlaşılması için genellikle aynı müşteriye birden fazla telefon bağlantısı gerekmektedir. Verilerde 16 adet giriş ve 1 adet çıkış verisi bulunurken sınıflandırma için etkili olmayandört adet giriş verisi kullanılmamıştır. Ayrıca sınıflandırma için sayısal veriler (age, campaign, previous) kategorilere ayrılarak nominal hale getirilmiştir.

Giriş verileri;

- age: yaş (18-24, 25-34, 35-44, 45-54, 55-64, 65-74, 75-90)
- job: meslek (yönetici, mavi yaka, girişimci, ev hanımı, yönetim, emekli, serbest meslek, hizmet sektörü, öğrenci, teknisyen, işsiz, bilinmiyor)
- martial: medeni hal (evli, bekâr, dul)
- education: eğitim (bilinmiyor, ilköğretim, ortaöğretim, üniversite)
- default: hâlihazırdakredisi var mıdır? (evet, hayır)
- balance: avro cinsinden yıllık ortalama bakiye (negatif, 0-100, 101-500, 501-1000, 1001-10000, >10000)
- housing: konut kredisi var mıdır? (evet, hayır)
- loan: bireysel kredi var mıdır? (evet, hayır)
- contact: iletişim temas türü (bilinmiyor, sabit telefon, cep telefonu)
- campaign: müşteriye mevcut kampanya boyunca yapılan temas sayısı (1, 2-4, 5-10, >10)
- previous: müşteriye mevcut kampanyadan önce yapılan temas sayısı (yok, 1, 2-5, >5)
- poutcome: bir önceki pazarlama kampanyasının sonucu (bilinmiyor, başarılı, başarısız, diğer)

Çıkış verisi;

- DEPOSIT: müşteriye vadeli mevduat verildi mi? (evet, hayır)

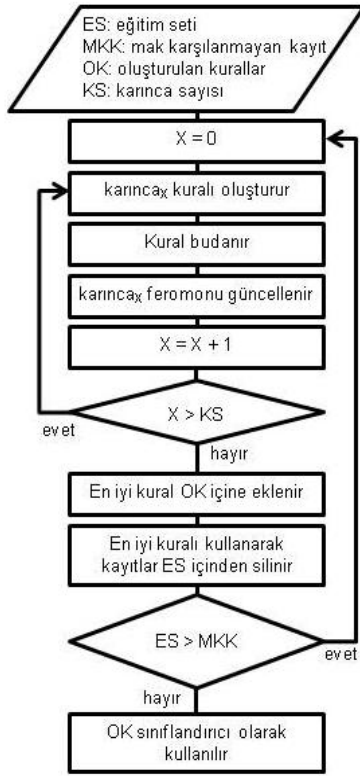
4.1. Karınca Kolonisi Optimizasyonu Sınıflandırma Algoritması

Bu çalışma kapsamında doğrudan bankacılık kampanyası analizinde KKOSA kullanılmış ve çok bilinen diğer algoritmalarla performansı kıyaslanmıştır. Karınca kolonisi optimizasyon algoritması (Dorigo & Stützle, 2004) ilk olarak Marco Dorigo tarafından doktora tezinde (Dorigo, 1992) tanıtılmıştır. Karıncaların yiyecek arayışı sürü işbirliğine dayanır. Yiyecek arayan her karınca hedefine doğru giderken ve yuvaya dönerken yol üzerinde feromon denilen bir salgı bırakır. Bu salgı geriden gelen karıncalara yol gösteren bir koku verir. Feromon zamanla yok olan bir salğıdır. Bu sebeple en kısa yoldan giden karıncalar daha fazla feromon bırakırlar. Arkadan gelen karıncalarda kokunun fazla olduğu bu kısa yolu seçerler. Böylece karınca sürüsü en kısa yoldan yiyeceğe ulaşır.

KKOSA özet olarak "eğer-öyleyse-sonuç" mantığına dayalı sınıflandırma kurallarının keşfedilmesidir. Karınca kolonisi ile kural tabanlı sınıflandırma (Parpinelli vd., 2002a) için ilk uygulama Parpinelli ve arkadaşları (Parpinelli vd., 2002b) tarafından geliştirilmiştir. Algoritma böl ve keşfet mantığına dayanır. Veriler öncelikle eğitim seti için ayrılır. Verilerin bilgi kazancı hesaplanır. Her bir karınca bilgi kazancı ve feromon seviyeleri baz alınarak hesaplanan bir olasılık formülüne göre kuralları oluşturmaya başlar. Her kuraldan sonra karıncaların feromon seviyeleri güncellenir. Böylece yeni bir kural keşfinde olasılık durumları değiştirilerek daha iyi bir kural bulmaya çalışılır. KKOSA ile sınıflama kuralları çıkarımı Şekil 2 ile gösterilmiş ve algoritma sözde kodları aşağıda verilmiştir.

Karınca kolonisi optimizasyonu sınıflandırma algoritması:

```
EğitimSeti [ ] /* Bütün eğitim verileri */
KeşfedilmişKurallar = [ ]; /* boş liste olarak başlatılır */
WHILE (EğitimSeti > KarşılanmayanMaksimumKayıtlar)
t = 1; /* karınca indeksi */
j = 1; /* yakınsama test indeksi */
/* Bütün izleri aynı oranda feromon ile ayarla */
REPEAT
Antı boş bir kural ile başlar her adımda yeni bir terimi ekleyerek sınıflandırma kuralı Rt oluşturur.
Rt budandır
Rt kuralının kalitesiyle orantılı olarak Antı takip edilen izlerin feromon seviyesi güncellenir
IF (Rt eşitse Rt-1) /* yakınsama testi güncellenir */
THEN j = j + 1;
ELSE j = 1;
END IF
t = t + 1;
UNTIL (i " Karınca sayısı) OR (j " Yakınsanan kural sayısı)
Rt içinden en iyi kuralı Reniyi seç.
En iyi kural Reniyi KeşfedilmişKurallar[ ] listesine eklenir
EğitimSeti içinden en iyi kural Reniyi tarafından doğru olarak kapsanan örnekler çıkarılır.
END WHILE
```



Şekil 2. KKSOA ile kural çıkarımı

Kurala eklenecek terimij olasılığı denklem 1 ile hesaplanır.

$$(1) P_{ij} = \frac{\eta_{ij} \cdot \tau_{ij}(t)}{\sum_{i=1}^a x_i \cdot \sum_{j=1}^{b_i} (\eta_{ij} \cdot \tau_{ij}(t))}$$

η_{ij} : terimij sezgisel değeri

$\tau_{ij}(t)$: t aşamasındaki i,j pozisyonundaki terimij ile alakalı feromon seviyesi

a: toplam nitelik sayısı

x_i : A_i niteliği karınca tarafından kullanılmadıysa 1, değilse 0 değeri alır

b_i : i-inci niteliğin toplam değer sayısı

Sezgisel değer (η_{ij}) terimij için kaliteyi hesaplar. Bu hesaplama denklem 2 ile verilen formülle bilgi kazancı teorisine göre entropiye dayalı bir ölçüm yapar.

$$(2) H(W|A_i = V_{ij}) = - \sum_{w=1}^k (P(w|A_i = V_{ij}) \cdot \log_2 P(w|A_i = V_{ij}))$$

w: tahmin edilecek sınıfın niteliği

k: sınıf sayısı

A_i : i-inci sıradaki nitelik

V_{ij} : A_i niteliğine ait j-inci sıradaki değeri

$P(w|A_i=V_{ij})$: İzlenen $A_i=V_{ij}$ koşuluna dayalı gözlemlenen w sınıfının olasılık değeri

$H(W|A_i=V_{ij})$ değeri 0 ile $\log_2 k$ arasında değiştiği için sezgisel fonksiyon denklem 3 ile verilen formüle göre normal hale getirilir.

$$(3) \eta_{ij} = \frac{\log_2 k - H(w|A_i=V_{ij})}{\sum_{i=1}^a x_i \cdot \sum_{j=1}^{b_i} (\log_2 k - H(w|A_i=V_{ij}))}$$

a: toplam nitelik sayısı

x_i : A_i niteliği karınca tarafından kullanılmadıysa 1, değilse 0 değeri alır

b_i : i-inci niteliğin toplam değer sayısı

Denklem 4 ile gösterilen formülde verilen feromon seviyesi tüm niteliklerin değerlerinin sayıları ile ters orantılıdır.

$$(4) \tau_{ij}(t = 0) = \frac{1}{\sum_{i=1}^a b_i}$$

Terim ij için feromon ayarlaması denklem 5 ile belirtilen formüle göre yapılır.

$$(5) \tau_{ij}(t + 1) = \tau_{ij}(t) + \tau_{ij}(t) \cdot Q, \forall i, j \in R$$

Kuralın kalitesi ise denklem 6 ile verilen formüle göre hesaplanır.

$$(6) Q = \text{hassasiyet} \cdot \text{kesinlik} = \frac{DP}{DP + YN} \cdot \frac{DN}{YP + DN}$$

DP (doğru pozitif), gerçekte kredi verilen ve algoritmanın da kredi verilebilir tahmin ettiği terimler
DN (doğru negatif), gerçekte kredi verilmeyen ve algoritmanın da kredi verilemez tahmin ettiği terimler
YP (yanlış pozitif), gerçekte kredi verilen ama algoritmanın kredi verilemez tahmin ettiği terimler
YN (yanlış negatif), gerçekte kredi verilmeyen ama algoritmanın kredi verilebilir tahmin ettiği terimler
Hassasiyet DP/(DP+YN) oranı
Kesinlik DN/(DN+YP) oranı

5. BULGULAR VE ÇIKARIMLAR

Verilerin analizi için Waikato Üniversitesi açık kaynak kodlu bir proje olan WEKA (Hall vd., 2009) makine öğrenmesi veri madenciliği yazılımı kullanılmıştır. Bu çalışmada, 10 katlı çapraz doğrulama tekniğine(Kohavi, 1995) göre, eğitim ve test kümelerine ayrılan veriler üzerinde KKOSA ve çok bilinen diğer algoritmalaragöre ilk önce %10'luk 4521 adet veri analiz edilmiş, elde edilen sonuçlar tablo 1'de gösterildiği gibi doğruluk oranları ve sınıflandırma modelinin oluşum zamanına göre kıyaslanmıştır.

Tablo 1. Algoritmaların %10'luk 4521 adet veri üzerinde sınıflandırma sonuçları

Algoritma	Doğru Sınıflandırma	Yanlış Sınıflandırma	Doğruluk Yüzdesi	Hata Yüzdesi	Model Oluşum Zamanı
Naive Bayes Yöntemi	3959	562	87.5691	12.4309	0.05 saniye
Logistic Fonksiyon	4029	492	89.1176	10.8825	1.97 saniye
J48 (C4.5) Decision Tree	4035	486	89.2502	10.7498	0.20 saniye
JRip Kuralları	4039	482	89.3306	10.6614	1.01 saniye
Karınca Kolonisi	4043	478	89.4271	10.5729	3.44 saniye

Elde edilen sonuçlara göre KKOSA kullanılarak, 3.44 saniye model oluşum süresinde 4521 kişinin %89.4271 doğruluk oranı ile kredi verilebilirlik ihtimallerin test edilmiştir. Ancak diğer algoritmalarla kıyaslandığında, sınıflandırma modelinin oluşum zamanı daha kısa olan algoritmalar mevcuttur. Örneğin, çok bilinen Naive Bayes algoritması 0.05 saniyede %87.57 doğruluk oranı yakalamıştır. Ayrıca sonuç tablosundan anlaşılacağı üzere algoritmaların doğruluk oranlarının birbirine yakın ancak model oluşum sürelerinin çok farklı olduğu görülmektedir. Model oluşum zamanı diğerlerine göre fazla olsa da doğruluk açısından KKOSA daha iyi olduğu anlaşılmaktadır. Aynı yöntem ile 45211 adet tüm veriler analiz edilerek sonuçlar tablo 2 ile gösterilmiştir.

Tablo 2. Algoritmaların 45211 adet veri üzerinde sınıflandırma sonuçları

Algoritma	Doğru Sınıflandırma	Yanlış Sınıflandırma	Doğruluk Yüzdesi	Hata Yüzdesi	Model Oluşum Zamanı
Naive Bayes Yöntemi	39744	5467	87.9078	12.0922	0.16 saniye
Logistic Fonksiyon	40356	4855	89.2615	10.7385	16.36 saniye
J48 (C4.5) Decision Tree	40387	4824	89.3300	10.6700	1.36 saniye
JRip Kuralları	40356	4855	89.2615	10.7385	3.96 saniye
Karınca Kolonisi	40341	4870	89.2283	10.7717	125.008 saniye

Sonuçlara göre, tüm verilerin analizinde KKOSA yönteminin başarı oranı düşmüştür. Verilerin artması ile karar ağaçları daha etkili bir sınıflama yapmıştır. Ayrıca KKOSA model oluşum zamanı da fazla bir oranda artmıştır.

Bir sınıflandırma testinin yeterliliğini değerlendirmek için alıcı işlem karakteristiği (Receiver Operating Characteristics) ROC analizi (Gribskov & Robinson, 1996) yapılarak %10'luk 4521 adet veri için tablo 3'de ve 45211 adet tüm veriler için tablo 4'de gösterilen durumlara göre hassasiyet ve kesinlik ölçütleri hesaplanmıştır. Sınıflandırma işleminde metotlar, doğru pozitif değerleri (hassasiyet) tespit etme ve yanlış pozitif değerleri (kesinlik) eleme kabiliyeti arasındaki dengeyi kurmakla uğraşırlar.

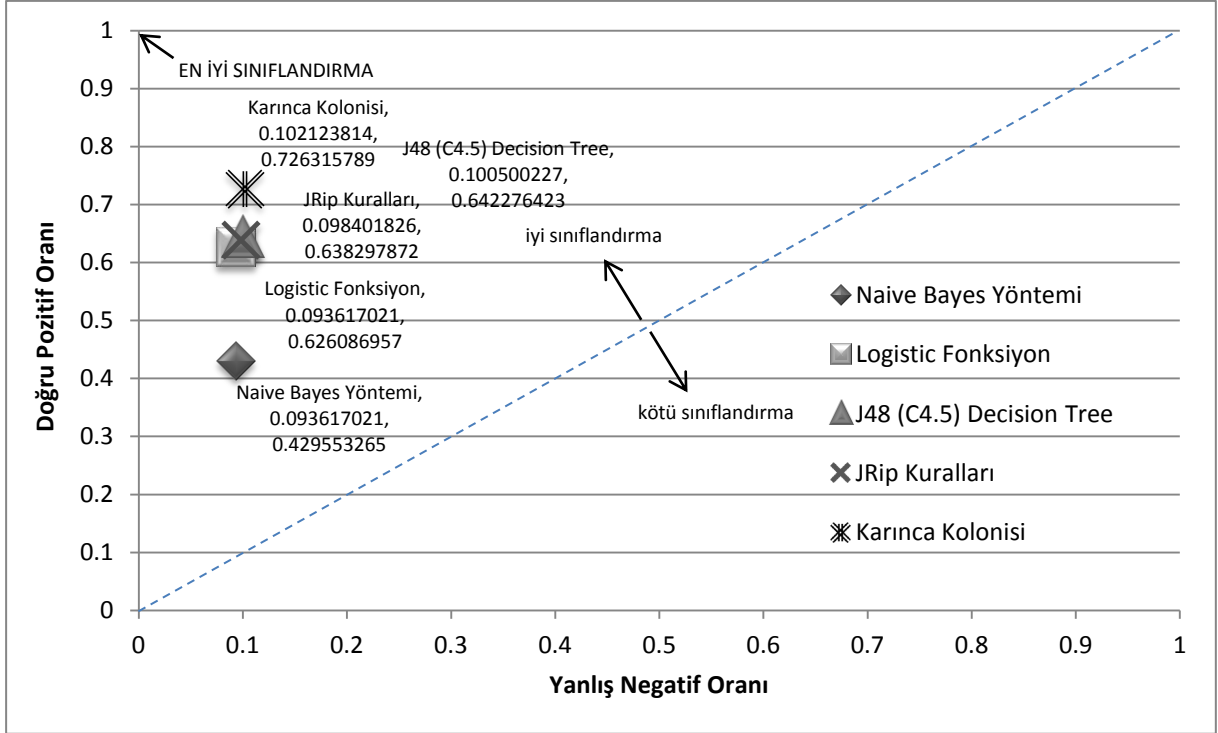
Tablo 3. %10'luk 4521 adet veri için ROC analizi sonuçları

Karar Ağacı	DP	DN	YP	YN	Hassasiyet	Kesinlik	Yanlış Negatif Oranı
Naive Bayes Yöntemi	125	3834	396	166	0,429553265	0,906382979	0,093617021
Logistic Fonksiyon	72	3957	449	43	0,626086957	0,898093509	0,101906491
J48 (C4.5) Decision Tree	79	3956	442	44	0,642276423	0,899499773	0,100500227
JRip Kuralları	90	3949	431	51	0,638297872	0,901598174	0,098401826
Karınca Kolonisi	69	3974	452	26	0,726315789	0,897876186	0,102123814

Tablo 4. 45211 adet tüm veriler için ROC analizi sonuçları

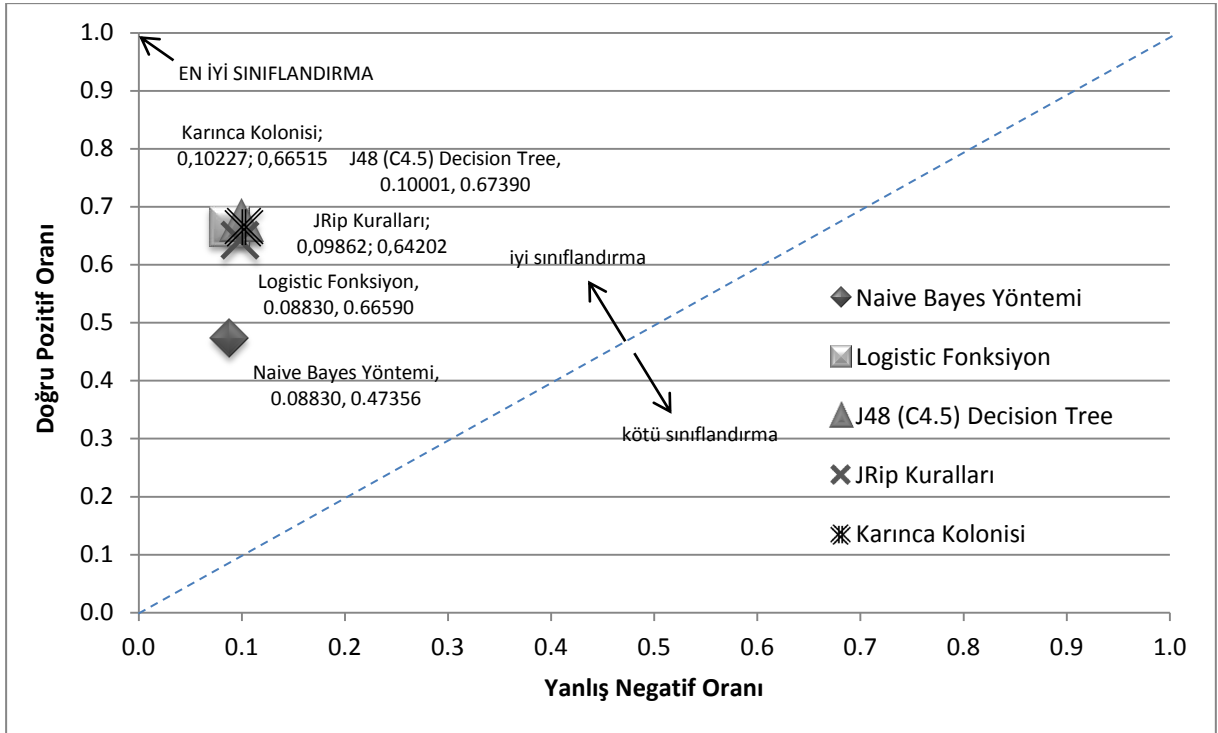
Algoritma	DP	DN	YP	YN	Hassasiyet	Kesinlik	Yanlış Negatif Oranı
Naive Bayes Yöntemi	1594	38150	3695	1772	0,473559121	0,911697933	0,088302067
Logistic Fonksiyon	871	39485	4418	437	0,665902141	0,899369064	0,100630936
J48 (C4.5) Decision Tree	901	39486	4388	436	0,673896784	0,899986324	0,100013676
JRip Kuralları	981	39375	4308	547	0,642015707	0,901380400	0,098619600
Karınca Kolonisi	729	39555	4506	367	0,665145985	0,897732689	0,102267311

Sınıflandırma işlemlerinde kesinlik ve hassasiyet arasındaki dengeyi değerlendirmek için ROC eğrisi x ekseninde yanlış negatif oranı değerleri ile y ekseninde doğru pozitif oranı değerleri çizilir ve altında kalan alan kullanılır. ROC puanı 1'e yaklaştıkça pozitifler daha iyi bir şekilde negatiflerden ayrılır. Şekil 3'de görüldüğü gibi KKOSA en iyi ROC sonucuna sahiptir.



Şekil 3. %10'luk 4521 adet veri için ROC analizi sonuç grafiği

45211 adet tüm veriler için hazırlanan ve şekil 4 ile gösterilen ROC analizinde %10'luk veri seti ile arasında küçük farklılıklar olsa da aynı şekilde KKOSA yöntemi en iyi sonuca sahiptir.



Şekil 4. 45211 adet tüm veri seti için ROC analizi sonuç grafiği

Aynı veri setinin kullanıldığı diğer bir çalışmada(Moro vd., 2011) toplam 45211 adet kayıt üzerinde Naive Bayes yöntemi, karar ağacı ve karar destek sistemleri test edilmiştir. Elde edilen sonuçlara göre karar destek sistemleri 0.938 ROC puanı ile en iyi sonucu elde etmiştir.

Telefon bankacılığında hedef müşteriye en hızlı ve en etkin yöntemlerle ulaşmak amaçlandığından bu çalışmada verilerin öncelikle %10'luk kısmı analiz edilmiştir. Ardından verilerin bütünü analiz edilmiştir. Verilerin bölümlendirilerek analiz yapılması için KKOSA uygun iken verilerin bütünü ele alındığında KKOSA performansında az da olsa düşme eğilimi gözlemlenmiştir.

6. SONUÇ

Elde edilen sonuçlara göre %10'luk 4521 adet veri için KKOSA diğer yöntemlere göre en iyi doğruluk oranı ve ROC grafiği göstermiştir. Ancak tüm veriler ele alındığında başarı oranı düşmüştür. Algoritmanın performansını artırmak için karınca sayısını arttırmak bir çözüm olabilir de model oluşum zamanı açısından süresi çok uzayacaktır. Elbette KKOSA kolay anlaşılır ve çabuk uygulanabilir bir algoritma olarak birçok bilimsel çalışmada rahatça kullanılabilir. Fakat veri sayısı arttıkça sezgisel bir algoritma olan KKOSA karar kuralları çıkarmada çok fazla zaman kaybetmekte ve başarı oranı düşmektedir.

Yapılan bu çalışmada, bankacılık sektöründe KKOSA yönteminin; pazarlama, ticari analiz, portföy yönetimi, risk yönetimi, müşteri ilişkileri yönetimi, müşteri kazanımı ve elde tutma, müşteri profili analizi ve müşteri hizmetleri alanında iş performansını artırmak amacı ile diğer yaygın yöntemler gibi iyi bir performansla kullanılabileceği tespit edilmiştir. Bu bağlamda diğer veri madenciliği yöntemleri gibi karınca kolonisi yöntemi de bankacılık veya bir başka sektör için yeni iş fırsatlarını; daha önce bilinmeyen desen keşfini sağlayarak ve tüketim eğilimleri ile müşteri davranışlarını analiz ederek iyi bir şekilde sunabilir.

KKOSA sadece nominal veriler ile çalışabilen bir algoritma sebebiyle sayısal değerler örneğin mevduat miktarı gibi nitelikler gruplandırılarak işlem yapılmıştır. Bu durum elbette performansı etkileyecektir. Daha farklı gruplamalar ile farklı doğruluk oranları elde edilebilir. Ayrıca ileriye dönük çalışmalar için örneğin karar ağaçları veya yapay sinir ağları gibi diğer algoritmalar ile KKOSA melezleştirilmesi yapılarak performans artışı sağlanabilir.

KAYNAKÇA

Ali, Ö. G., & Artürk, U. (2014). Dynamic churn prediction framework with more effective use of rare event data: The case of private banking. *Expert Systems with Applications*, 7789-7903.

Argüden, Y., & Erşahin, B. (2008). *Veri Madenciliği Veriden Bilgiye Masrafdan Değere*. İstanbul: ARGE Danışmanlık A.Ş.

Bach, M. P., Juković, S., Dumići, K., & Šarlija, N. (2013). Business Client Segmentation in Banking Using Self-Organizing Maps. *South East European Journal of Economics and Business*, 32-41. doi:10.2478/jeb-2013-0007

Bramer, M. (2013). *Principles of Data Mining*. London: Springer-Verlag.

- Brito, P. Q., Soares, C., Almeida, S., Monte, A., & Byvoet, M. (2015). Customer segmentation in a large database of an online customized fashion business. *Robotics and Computer-Integrated Manufacturing*, 1-8.
- Crone, S. F., Lessmann, S., & Stahlbock, R. (2006). The impact of preprocessing on data mining: An evaluation of classifier sensitivity in direct marketing. *European Journal of Operational Research*, 781-800.
- Domingo, R. T. (2015, 01 28). Applying Data Mining to Banking. The one stop business management knowledge center: <http://www.rtdonline.com/BMA/BSM/4.html> adresinden alınmıştır
- Dorigo, M. (1992). *Optimization, Learning and Natural Algorithms*. PhD thesis. Milano, Italy: Politecnico di Milano.
- Dorigo, M., & Stützle, T. (2004). *Ant Colony Optimization*. London: MIT Press.
- Elsalamony, H. A. (2014). Bank Direct Marketing Analysis of Data Mining Techniques. *International Journal of Computer Applications* , 12-22.
- European Central Bank. (2012). *Report on Card Fraud*. Frankfurt: European Central Bank.
- Govindarajan, M. (2014). A Hybrid RBF-SVM Ensemble Approach for Data Mining Applications. *I.J. Intelligent Systems and Applications*, 84-95.
- Gray, G. L., & Debreceeny, R. S. (2014). A taxonomy to guide research on the application of data mining to fraud detection in financial statement audits. *International Journal of Accounting Information Systems*, 357-380.
- Gribskov, M., & Robinson, N. L. (1996). Use of receiver operating characteristic (ROC) analysis to evaluate sequence. *Computers & Chemistry*, 25-33.
- Gürsoy, U. T. (2009). *Veri Madenciliği ve Bilgi Keşfi*. Ankara: Pegem Akademi.
- Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., & Witten, I. H. (2009). The WEKA data mining software: an update. *SIGKDD Explor. Newsl.*, 11(1), 10-18.
- Han, J., & Kamber, M. (2006). *Data Mining: Concepts and Techniques*. San Francisco: Morgan Kaufmann Publishers.
- Han, J., Kamber, M., & Pei, J. (2006). *Data Mining, Southeast Asia Edition: Concepts and Techniques*. San Francisco: Morgan Kaufmann Publishers.
- Hany, A. E., & Alaa, M. E. (2013). Bank Direct Marketing Based on Neural Network. *International Journal of Engineering and Advanced Technology (IJEAT)*, 392-400.
- Javaheri, S. H., Sepehri, M. M., & Teimourpour, B. (2014). Response Modeling in Direct Marketing: A Data Mining-Based Approach for Target Selection. Y. Zhao, & Y. Cen içinde, *Data Mining Applications with R* (s. 153-180). Waltham: Elsevier Inc.

- Keramati, A., Jafari-Marandi, R., Aliannejadi, M., Ahmadian, I., Mozaffari, M., & Abbasi, U. (2014). Improved churn prediction in telecommunication industry using data techniques. *Applied Soft Computing*, 994-1012.
- Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. *IJCAI'95 Proceedings of the 14th international joint conference on Artificial intelligence* (s. 1137-1143). Montreal, Quebec, Canada: Morgan Kaufmann Publishers Inc. San Francisco, CA, USA ©1995 .
- Kosina, P., & Gama, J. (2015). Very fast decision rules for classification in data streams. *Data Mining and Knowledge Discovery*, 168-202.
- Köktürk, M. S., & Dirsehan, T. (2012). *Veri Madenciliği ile Pazarlama Etkileşimi* . Ankara: Nobel Akademik Yayıncılık.
- Kumari, B., & Shrivastava, V. (2015). Evaluation and Comparison of Performance of different Classifiers. *International Journal of Emerging Trend in Engineering and Basic Sciences (IJEEBS)*, 604-611.
- Kuo, R., Huang, Y., Chih-Chieh, L., Yung-Hung, W., & Ferani, Z. E. (2014). Automatic kernel clustering with bee colony optimization algorithm. *Information Sciences*, 107-122.
- Liao, S.-h., Chen, Y.-j., & Hsieh, H.-h. (2011). Mining customer knowledge for direct selling and marketing. *Expert Systems with Applications*, 6059–6069.
- Liébana-Cabanillasa, F., Noguerasb, R., Herrerac, L., & Guillén, A. (2013). Analysing user trust in electronic banking using data mining methods. *Expert Systems with Applications*, 5439–5447. doi:10.1016/j.eswa.2013.03.010
- Moro, S., Laureano, R., & Cortez, P. (2011). Using Data Mining for Bank Direct Marketing: An Application of the CRISP-DM Methodology. *Proceedings of the European Simulation and Modelling Conference - ESM'2011 October* (s. 117-121). Guimarães, Portugal: EUROSIS. <https://archive.ics.uci.edu/ml/datasets/Bank+Marketing#> adresinden alınmıştır
- Nachev, A. (2014). *Application of Data Mining Techniques for Direct Marketing*. Sofia: Ithea.
- Oz, E. (2006). *Management Information Systems*. Canada: Thomson Course Technology.
- Özkan, Y. (2013). *Veri Madenciliği Yöntemleri*. İstanbul: Papatya Yayıncılık .
- Parpinelli, S. R., Heitor, S. L., & Alex, F. A. (2002a). An ant colony algorithm for classification rule discovery. H. A. Abbass, R. A. Sarker, & C. S. Newton içinde, *Data Mining: A Heuristic Approach* (s. 191-208). Idea Group Pub.
- Parpinelli, R. S., Lopes, H. S., & Freitas, A. A. (2002b). Data mining with an ant colony optimization algorithm. *Evolutionary Computation, IEEE Transactions on*, 321-332.
- Popelka, O., Hřebíček, J., Štencl, M., Hodinka, M., & Trenz, O. (2012). Comparison of different non-statistical classification methods. *30th International Conference*

- Mathematical Methods in Economics (s. 727-732). Karvina: Silesian University in Opava.
- Ravisankara, P., Ravi, V., Rao, G. R., & Bose, I. (2011). Detection of financial statement fraud and feature selection using data mining techniques. *Decision Support Systems*, 491-500.
- Shaw, M. J., Subramaniama, C., Tan, G. W., & Welge, M. E. (2001). Knowledge management and data mining for marketing. *Decision Support Systems*, 127-137.
- Silahtaroglu, G. (2013). *Veri Madenciliği Kavram ve Algoritmalar*. İstanbul: Papatya Yayıncılık Eğitim.
- Sumathi, S., & Sivanandam, S. (2006). *Introduction to Data Mining and Its Applications*. Tamil Nadu: Springer Science & Business Media.
- Wijaya, T. K. (2013). From Data to Knowledge: Data Mining. IRMA içinde, *Data Mining: Concepts, Methodologies, Tools, and Applications* (s. 1150-1162). Hershey: Information Resources Management Association.