

# Makine Öğrenme Yöntemleri ile Kredi Risk Analizi

## Credit Risk Analysis Using Machine Learning Algorithms

Sacide Kalaycı ve Mustafa Kamaşak  
Bilgisayar Mühendisliği  
İstanbul Teknik Üniversitesi  
İstanbul, Türkiye  
{kalaycisa,kamasak}@itu.edu.tr

Seçil Arslan  
Ar-Ge ve Özel Projeler  
YapıKredi Teknoloji  
İstanbul, Türkiye  
secil.arslan@ykteknoloji.com.tr

**Özetçe** —Müşteri kredi riski analizinde, müşterinin kredi başvuru riskini değerlendirmenin yanı sıra, devam eden bir kredinin de riski değerlendirilerek karar almak, finansal merkezlerin risklerini azaltmak için çok önemlidir. Türkiye’de yapılan son çalışmalar göstermiştir ki bireysel, kurumsal ve KOBİ gibi müşterilere sağlanan krediler arasında, KOBİ kredileri finansal merkezler için en riskli krediler grubuna girmektedir. Bu nedenle bu çalışmada KOBİ müşterilerin kredilerinde sergilediği geri ödeme davranışları ile birlikte müşteri bilgileri de kullanılarak müşterinin 6 aylık bir dönemdeki performansı tahmin edilmeye çalışılmıştır. Rastgele orman algoritması, sinir ağları, destek vektör makineleri, meylli hızlandırma yöntemleri kullanılarak performans karşılaştırılması yapılmış ve ayrıca kullanılan ödeme davranışı özniteliklerinin problem üzerinde etkisini analiz edebilmek için testler gerçekleştirilmiştir. Son olarak yapılan deneylerle, harmanlanmış öğrenme tekniklerini kullanmanın sistemin performansı üzerinde olumlu bir etki yarattığı gözlemlenmiştir.

**Anahtar Kelimeler**—KOBİ müşteri risk analizi, harmanlanmış öğrenme yöntemleri, rastgele orman algoritması, sinir ağları, destek vektör makineleri, meylli hızlandırma

**Abstract**—In credit risk analysis, besides assessing risk of credit applications, taking decision by foreseeing risk of active credit is very important to decrease risk of financial institutions. In Turkey, recent studies reveal that for financial institutions, risk of SME credits is higher than other credit types such as consumer and corporate. Therefore, this paper focuses on predicting SME customer status for period of six months by utilizing application scoring additional to customer behaviour features. By utilizing Random Forest, Neural Networks, Support Vector Machines and Gradient Boosting, performance comparison and also feature analysis for customer behaviour are conducted. Finally, conducted experiments show that utilizing Stacked Generalization methods has positive effect on performance of SME credit risk analysis.

**Keywords**—SME credit risk analysis, stacked generalization, random forest, neural network, support vector machines, gradient boosting

### I. GİRİŞ

Bankacılık Düzenleme ve Denetleme Kurumu’nun (BDDK) yayınladığı rapora göre Türkiye’de küçük ve orta

büyükteki işletme (KOBİ) durumundaki müşterilerin kullandıkları kredi oranları yıllar geçtikçe artmakta ve bu nedenle de finansal merkezler için müşteri riskinin daha iyi tahmin edilebilmesi gittikçe önem kazanmaktadır. Bu ihtiyaçtan dolayı son yıllarda müşteri riskini daha iyi analiz edebilmek için uzman bilgi ve tecrübesine dayalı klasik yöntemler yerine makine öğrenmesi kullanılarak geliştirilen sistemler kullanılmaktadır [1]–[3]. Müşteri kredi riski analizi; müşterinin gelecek bir tarihte finansal bir dar boğaz yaşayıp yaşamayacağını tahmin ederek, finansal merkezin riskini en aza indirmeyi amaçlar. Thomas ve diğ. [4] tarafından yapılan tanımlamalara göre müşteri riski analizi ikiye ayrılır: Müşteri başvuru skor analizi ve müşterinin ödeme davranışının analizi. Müşteri başvuru skor analizinde müşterinin kredi başvurusu sırasında sunduğu ya da geçmiş bilgileri kullanılarak müşterinin ileri bir zamandaki durumu tahmin edilmeye çalışılır. Müşteri ödeme davranışı analizinde ise müşterilerin krediye başvurduktan sonraki belirli bir süre boyunca ödeme davranışları incelenerek müşterinin durumu tahmin edilmeye çalışılır. Müşteri ödeme davranışını incelemek için yapılan çalışmaların bazıları müşterinin herhangi bir zamanda finansal bir sıkıntı yaşayıp yaşamayacağı konusuna odaklanırken, diğerleri de müşterinin ne zaman ya da herhangi bir süre aralığında sıkıntı yaşayıp yaşamayacağına odaklanır.

Türkiye’de KOBİ’ler için bireysel müşterilerden farklı olarak tek tipte bir kredi (ihtiyaç kredisi, ev kredisi vb.) ya da ürün (kredi kartı, esnek hesap) teklif etmek yerine belirli bir limitte, içinde birçok farklı ürün barındıran ve belirli bir süre geçerliliği olan bir teklif paketi oluşturulur. KOBİ’ler için hazırlanan her bir teklif paketi KOBİ’lerin ihtiyaçlarına göre hazırlanmış birbirinden farklı ödeme planları bulunan 77 farklı üründen (çek defteri, kredi kartı vb.) birkaçını içerebilir ve toplam limit bu ürünler arasında paylaştırılır. KOBİ müşterileri, bu teklif paketinin geçerli olduğu süre boyunca ihtiyacı olan kredi ya da ürünlerden istediğini dilediği zaman kullanabilir. Eğer müşteri kullandığı ürünün ödemesini 90 gün boyunca gerçekleştirmezse kredinin durumu “problemli kredi” (non-performing loan (NPL)) olarak belirlenir. NPL durumundaki herhangi bir ürünün sahibi de “NPL durumundaki müşteri” olarak nitelendirilir. Aksi takdirde kredi ve ilgili müşterinin durumu “sağlıklı kredi” (performing loan (PL)) olarak devam eder.

Finansal merkezler için müşterinin ne zaman finansal sıkıntı yaşacacağı önemli olduğundan bu çalışma kapsamında KOBİ müşterilerinin 6 aylık bir zaman diliminde finansal sıkıntı yaşayıp yaşamayacağına odaklanılmıştır. Bu motivasyonla Türkiye'deki KOBİ müşterilerinin 1 Ocak 2015 - 1 Ekim 2016 tarihleri arasındaki verileri analiz edilerek 1 Nisan 2016 - 1 Ekim 2016 tarihleri arasında NPL olup olmayacakları makine öğrenmesi yöntemleri kullanılarak tahmin edilmeye çalışılmıştır. Bu amaçla, müşteri ödeme davranışı probleminde ve müşterilerin başvurularını değerlendirmek için kullanılan öznitelikler kullanılmıştır.

Kim ve diğ. [5] Kore'deki KOBİ müşterilerinin riskini analiz etmek için destek vektör makinelerini (Support Vector Machine (SVM)) kullanmış ve SVM'in, sinir ağları (Neural Network (NN)) ve lojistik regresyon (Logistic Regression (LR)) yöntemlerinden daha başarılı olduğunu göstermişlerdir. Thomas ve diğ. [6] müşterilerin tüketici kredilerindeki durumunu tahmin etmek için LR ve sağkalım analizi (survival analysis (SA)) kullanmışlar ve yaptıkları deneylerle ilk bir yılda NPL olacak müşteriler için SA yönteminin daha başarılı olduğunu kanıtlamışlardır. Noh ve diğ. [7] NPL olan müşterileri tahmin etmek için SA yöntemi daha başarılıyken PL olan müşterileri tahmin etmek için LR ve NN yöntemlerinin daha başarılı olduklarını görmüşlerdir.

Çalışma ana hatlarıyla şu bölümleri içermektedir: 2. bölümde banka tarafından sağlanan KOBİ müşterilerinin 2015-2016 yılları arasındaki kredi davranışlarını içeren veri seti ve makine öğrenmesi için çıkarılan öznitelikler detaylı olarak anlatılmıştır. 3. bölümde yapılan deneyler ve elde edilen sonuçlar tartışılmış, 4. bölümde ise elde edilen sonuçlar özetlenmiştir.

## II. VERİ SETİ

Türkiye'de, Yapı Kredi Bankası (YKB) bünyesinde bulunan KOBİ müşterilerinin, 1 Ocak 2015 - 1 Ekim 2016 tarihleri arasındaki kredi verileri kullanılarak veri seti oluşturulmuştur. Müşterilerin tekliflerinin açılışından 1 Nisan 2016 tarihine kadar olan ödeme davranışları öznitelikleri çıkarmak için, 1 Nisan 2016 tarihinden sonraki davranışları ise müşterilerin durumunu (sınıfını) belirlemek için kullanılmıştır. Eğitilen bu sistem müşterilerin 1 Nisan - 1 Ekim tarihleri arasındaki geri ödeme performanslarının NPL ya da PL durumunda olacağını tahmin etmek için kullanılmıştır. Model geliştirmek için 1 Nisan 2016 tarihinden en az 6 ay önce açılmış ve bu tarih sonrasında da hala aktif ödemesi devam eden müşteriler veri setine dahil edilmiştir. PL durumunda olan müşteriler için tekliflerinin 1 Nisan - 1 Ekim tarihlerinde kapanmış olma koşulu aranmamıştır. Yani müşterileri PL olarak belirlemek için gerek ve yeter koşul olarak 1 Ekim 2016 tarihine kadar NPL durumuna düşmemiş olması aranmıştır. Projede müşterilerin 6 aylık süredeki davranışının tahmin edilmesi amaçlandığından, müşterilerin 1 Ekim 2016 tarihinden sonra NPL durumuna düşüp düşmemesi bu proje kapsamında müşterilerin sınıfını PL'den NPL'e değiştirmemiştir. Bu durum veri setimiz içinde benzer davranış gösteren fakat ödeme davranışı açısından farklı sınıflarda (PL ya da NPL) olan örnekler yol açabilecek olması göz önünde bulundurulmuş fakat bu örnekler veri setinden kaldırılmamıştır. Sistemin her ay çalışması ve zaman içerisinde kendini uyarlaması planlanmıştır.

KOBİ müşterilerinin risklerini belirlemek için çıkarılan öznitelikleri demografik bilgileri, başvuru bilgileri, şirketin

Tablo I: KOBİ MÜŞTERİLERİNİN EN ÇOK TERCİH ETTİĞİ ÜRÜNLER

Ürün Adı	Kullanım Oranı
Kredi Kartı	%20
Çek Defteri	%13
Ticari Kredi	%12
Diğer Ürünler	%55

finansal durumu, diğer bankalardaki riskleri, gösterilen teminat belgeleri ve ödeme davranışları olmak üzere 6 farklı kategoriye ayırmak mümkündür. Genellikle, KOBİ'lerin kurucuları aynı zamanda şirketin yöneticisi konumunda olduğu için büyük kuruluşların aksine KOBİ'ler, kurucularının kredi özellikleriyle de ilintilidir. Bu nedenle daha iyi eğitim almış, daha tecrübeli olan bir kurucu, karşısına çıkan olanakları daha iyi değerlendirebilir [8]. Bu nedenle müşterilerin demografik bilgileri için yaşı, eğitim durumu, cinsiyeti, maaşı gibi müşteri bilgileri öznitelik vektörüne eklenmiştir. Müşterilerin diğer bankalardaki riskleri için KKB (Kredi Kayıt Bürosu)'de sunulan bilgiler alınmış ve öznitelik vektörüne eklenmiştir. Firmanın finansal durumunu analiz etmek için müşterilerin bankaya verdikleri bilanço bilgileri kullanılmıştır. Firmanın kuruluş yılı, sektörü, ortaklık yapısı gibi özellikleri model içinde yansıtılabilmek için müşterilerin başvuru sırasında verdikleri bilgiler kullanılmıştır. Bunlara ek olarak, müşterilerin kredi başvuruları sırasında gösterdikleri teminatların türü, miktarı gibi bilgiler de öznitelik vektörüne eklenmiştir.

Müşterilerin ödeme davranışlarını analiz etmek içinse banka kapsamında bulunan 77 ürünün hepsine odaklanmak yerine, müşterilerin en çok kullanmayı tercih ettiği 3 ürünün detaylı analizi gerçekleştirilmiş ve müşterilerin bu ürünlerdeki davranışları incelenmiştir. Müşterilerin kullandıkları diğer ürünlerdeki davranışlarını da yansıtmak amacıyla kalan diğer ürünlerdeki genel davranışları incelenmiştir. Banka kapsamında müşterilerin en çok kullandığı ürünler Tablo I'de gösterilmiştir. Müşterilerin kredi kartı ödeme davranışlarını incelemek için gecikme faiz oranları, nakit çekme oranları, şüpheli hareketleri gibi bilgiler kullanılırken, çek defteri için müşterinin karşılıksız çıkan çek miktarları, gecikmeli olarak ödenen çek miktarları gibi bilgiler kullanılmıştır. Müşterilerin ticari kredisi içinse müşterilerin son 6 ayda geri ödemesindeki gecikme oranları incelenmiştir. Kalan diğer ürünleri temsil etmek için ise tüm ürünlerdeki toplam gecikme oranları, toplam faiz oranları gibi bilgiler öznitelik vektörüne eklenmiştir. Toplamda 3904 (1952 PL ve 1952 NPL) müşterinin 366 farklı özniteliği ile veri seti oluşturulmuş ve veri setinin %80'i eğitim için %20'si test için ayrılmıştır.

## III. DENEYSEL SONUÇLAR

Müşterilerin kredi risklerini tahmin etme probleminde kullanılan yöntemlerin ve özniteliklerin etkilerini analiz etmek için farklı test ortamları hazırlanmıştır. İlk olarak farklı makine öğrenme yöntemlerinin bu problem üzerinde başarısı incelenmiştir. Daha sonra, müşterilerin ödeme davranışlarının problem üzerindeki etkilerini analiz etmek için müşterilerin farklı ürünlerdeki ödeme davranışları veri setine eklenmiş ve sistemin başarısı üzerine olan etkisi incelenmiştir. Son olarak ise farklı yöntemlerin ürettiği çıktılar kullanılarak harmanlanmış öğrenme yöntemleri ile sistemin başarımı artırılmaya

çalışılmıştır [9]. Sistemin başarı ölçütü olarak müşteri riski analizi için yapılan çalışmalarda tercih edilen doğruluk ölçütü [3], [5] ve hata matrisleri (confusion matrix) kullanılmıştır.

#### A. Farklı Makine Öğrenmesi Yöntemlerinin Müşteri Riski Probleminde Başarımları

Bilgimiz dahilinde, bugüne kadar Türkiye’deki KOBİ müşterilerinin ödeme davranışları kullanılarak müşterinin gelecek 6 aydaki durumunu tahmin etmeye yönelik yapılmış bir çalışma bulunmamaktadır. Bu nedenle farklı yöntemlerin başarımını test etmek için, destek vektör makineleri (Support Vector Machines (SVM)) [10], çok katmanlı algılayıcılar (Multilayer Perceptron (MLP) - Neural Network (NN)), müşteri riski analizi problemi için yapılan çalışmalarda da sıklıkla kullanılan [2], [11], [12] rastgele orman algoritması (Random Forest (RF)) [13] ve son olarak da meyilli hızlandırma (Gradient Boosting (GB)) [14] yöntemleri kullanılmıştır.  $k(k = 5)$  katlı çapraz sorgulama yöntemi kullanılarak eğitim kümesi üzerinde ve sonrasında da oluşturulan modelin test kümesi üzerindeki başarımları hesaplanmıştır. Yöntemlerin  $k$  katlı çapraz sorgulama başarımlarının ortalamaları, standart sapmaları ve test kümesi üzerindeki başarımları Tablo II’de verilmiştir. Tablo II’den de görüldüğü gibi en iyi başarıyı GB algoritması, en iyi ikinci performansı ise rastgele orman algoritması vermiştir. Algoritmaların standart sapmalarını incelediğimiz de en düşük standart sapmaya NN sahipken ikinci en düşük standart sapmaya ise GB sahiptir.

Tablo II: MAKİNE ÖĞRENMESİ YÖNTEMLERİNİN BAŞARIMLARI

	K Katlamalı Çapraz Doğrulama Ortalaması	Test Kümesindeki Başarı Oranları
GB	% 84.20±0.9	%83.05
RF	% 83.08±1.5	%82.25
NN	% 80.10±0.6	%79.92
SVM	% 78.25±2.1	%79.41

#### B. Müşteri Ödeme Davranışı İle İlgili Özniteliklerin Müşteri Riski Problemine Etkisi

Bugüne kadar bu alanda yapılan çalışmaların birçoğunda müşterilerin kredi başlangıcındaki durağan (statik) öznitelikleri kullanılmış, belirli bir zaman aralığında olan ödeme davranışları kullanılmamıştır. Oysaki, müşterilerin ödeme davranışları müşterilerin ekonomik durumu hakkında bize daha doğru bilgiler verebilir ve daha iyi çalışan sistemler geliştirmemize olanak sağlayabilir [15]. Bu nedenle müşterilerin geri ödeme davranışlarının sistem üzerindeki etkisini incelemek için veri seti, müşterinin farklı ürünlerdeki davranışlarını tek tek ekleyerek genişletilmiş ve sistemin başarımına olan etkisi incelenmiştir. İlk olarak veri setinden müşterilerin ödeme davranışlarını içeren tüm öznitelikler çıkarılarak temel öznitelik vektörü oluşturulmuştur. Sonrasında ise sırasıyla müşterilerin ticari kredisindeki, çek defterindeki, kredi kartındaki ve diğer kalan ürünlerdeki davranışlarını içeren öznitelikler eklenmiştir. Öznitelik vektörüne müşterilerin farklı ürünlerdeki davranışları eklendikçe, özniteliklerin sistem üzerinde etkisi GB kullanılarak analiz edilmiştir. Sistemin  $k(k = 5)$  katlamalı çapraz sorgulama sonuçlarının ortalaması ve standart sapması ayrıca sistemin test kümesindeki başarıları

Tablo III’de gösterilmiştir. Tablo III’de de görüldüğü gibi ödeme davranışlarını incelemek, sadece sistemin başarısını artırmamış aynı zamanda modelin standart sapmasını da düşürerek, daha dengeli bir model elde edilmesine olanak sağlamıştır.

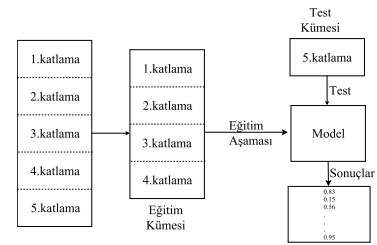
Tablo III: MÜŞTERİLERİN FARKLI ÜRÜNLERDEKİ ÖDEME DAVRANIŞLARININ SİSTEM PERFORMANSINA ETKİSİ

Ürün Tipi	Ortalama	Standart Sapma	Test Başarı
<b>Temel Özellikler</b>	<b>%76.65</b>	<b>%1.99</b>	<b>%76</b>
Ticari Kredi	+%0.80	-%0.09	+%0.75
Çek Defteri	+%1.87	-%0.41	+%1.75
Kredi Kartı	+%2.30	-%0.16	+%2.00
Diğer Ürünler	+%2.58	-%0.41	+%2.55
<b>Toplam</b>	<b>%84.20</b>	<b>%0.92</b>	<b>%83.05</b>

#### C. Müşteri Riski Probleminde Harmanlanmış Öğrenme Teknikleri

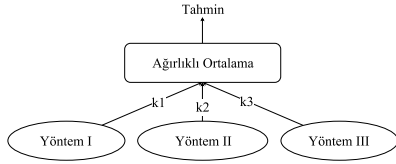
Son yıllarda sıklıkla kullanılan harmanlama yöntemleri ile başarılı sonuçlar elde edilmiştir [16]. Harmanlama yöntemlerinde, farklı makine öğrenme yöntemlerinin çıktıları ile bir ara katman oluşturulur ve bu ara katmandaki veriler öznitelik olarak kullanılarak bir sonraki katmana iletilir. Son katmanda ise bu öznitelikler kullanılarak tahminler elde edilmeye çalışılır. Bu sistemlerde istenildiği kadar katman oluşturulabilir fakat genel olarak iki katmanlı sistemler tercih edilmektedir.

Yapılan deneylerle, yöntemlerin, KOBİ müşterilerinin risklerini içeren veri seti üzerindeki başarımları incelenmiştir. Sonrasında ise bu yöntemlerin harmanlanmış şekilde kullanılarak elde edilebilecek olan başarımlarını incelemek için her yöntem Şekil 1’de gösterildiği gibi her  $k$ . kat için ( $k = 5$ ) eğitim kümesinin geri kalan örnekleri ile eğitilmiş ve  $k$ . kat için sınıf olasılıkları üretilmiştir. Her bir  $k$ . kat için yöntemler 20 farklı rastgele sayı ile kullanılarak eğitilmiş ve üretilen 20 farklı sınıf olasılıklarının ortalaması alınarak  $k$ . kat için yeni öznitelikler elde edilmiştir. Bu işlem her bir kat için yapılmış ve yeni eğitim kümesi oluşturulmuştur. Son olarak test kümesi için de aynı işlemler yapılarak test kümesinin sınıf olasılıkları ortalaması elde edilmiştir. Tüm bu işlemler seçilen her yöntem için tekrarlanmıştır.



Şekil 1: K. katlama için ilk katmanın oluşturulması

Veri seti üzerinde yapılan deneyler sonucunda en başarılı sonuçları üreten üç yöntemin GB, RF ve NN olduğu görülmüştür. Bu nedenle ilk olarak bu üç yöntem kullanılarak harmanlanmış bir sistem oluşturulmuş ve sistemin tasarımı Şekil 2’te gösterilmiştir. Bu 3 farklı yöntemden elde edilen veriler ilk olarak farklı makine öğrenmesi yöntemleri kullanılarak harmanlanmış fakat en başarılı yöntemin, Tablo II’de



Şekil 2: Üç farklı yöntem kullanarak harmanlanmış öğrenme sistemi

gösterilen  $k(k = 5)$  katlı çapraz doğrulama başarımlarına göre belirlenen katsayılarla çarpıldığında elde edildiği görülmüştür. Daha sonra ise GB, RF ve SVM kullanılarak harmanlanmış bir sistem oluşturulmuş ve son olarak da sadece en başarılı iki yöntem olan GB ve RF kullanılarak harmanlanmış bir sistem oluşturulmuştur. Elde edilen sonuçlar ve sistemde kullanılan yöntemlerin katsayıları Tablo IV’de gösterilmiştir.

Tablo IV: HARMANLANMIŞ ÖĞRENME SİSTEMİNİN SONUÇLARI

Yöntem Adı	Katsayılar	Başarımlar(%)	GB Başarısı(%)
GB, RF, NN	$k_1 = 0.3387, k_2 = 0.3354, k_3 = 0.3259$	%84.65	% <b>83.05</b>
GB, RF, SVM	$k_1 = 0.3394, k_2 = 0.3361, k_3 = 0.3245$	%85.29	% <b>83.05</b>
GB, RF	$k_1 = 0.5024, k_2 = 0.4976$	%85.93	% <b>83.05</b>

En başarılı sonuç RF ve GB yöntemlerini kullanarak oluşturulan harmanlanmış sistemle elde edilmiştir. Tablo IV’den de görüldüğü gibi NN ya da SVM kullanılarak oluşturulan sistemler en başarılı iki yöntem kullanılarak elde edilen başarımdan daha iyi bir sonuç sergileyememişlerdir. Son olarak GB ve RF kullanılarak elde edilen harmanlanmış yöntem ile GB yönteminin tek başına sergilediği başarımları karşılaştırmak için Tablo V’de hata matrisleri verilmiştir.

Tablo V: YÖNTEMLERİN HATA MATRİSLERİ

	Tahmin					
	Harmanlanmış Yöntem			GB		
	PL	NPL		PL	NPL	
Gerçek Durum	PL	%86.2	%13.8	PL	% 82.6	% 17.4
	NPL	%14.3	%85.7	NPL	% 16.5	% 83.5

Elde ettiğimiz sonuçları kullanıma açık veri seti üzerinde yapılan diğer çalışmalarla karşılaştırmaktansa Yapı Kredi kapsamında daha önceden yapılan bir çalışma ile karşılaştırmak daha anlamlı olacağı için, Derelioğlu ve Gürgeç [3] tarafından KOBİ müşterilerinin risk analizi için yapılan çalışmaları incelenmiştir. Derelioğlu ve Gürgeç, 512 KOBİ müşterisi için 27 öznetelik kullanarak müşteri riskini tahmin etmeyi amaçlamış ve en iyi sonuç olarak %76.17 başarımla elde edebilmişlerdir. Tablo IV’den görüldüğü gibi bu çalışma kapsamında oluşturulan sistem ile müşteri riskini analiz etme probleminde daha önceden yapılmış olan çalışmadan daha iyi bir performans elde edilmiştir.

#### IV. SONUÇ

Bu çalışmada, Türkiye’deki KOBİ müşterilerinin 2015-2016 yılları arasındaki kredi ödeme davranışları ve müşte-

rilerin bilgileri incelenerek 6 ay önceden müşterilerin kredi ödemelerinde sıkıntı yaşayıp yaşamayacaklarını tahmin etmeye dayalı bir sistem geliştirmiştir. Bilgimiz dahilinde Türkiye’de KOBİ müşterilerinin risklerini analiz etmek için müşteri bilgilerinin yanında, müşteri davranışları ilk kez kullanılarak, müşterinin 6 ay içindeki durumu tahmin edilmeye çalışılmıştır. İlk olarak müşteri riski probleminde sıklıkla kullanılan yöntemler tek başlarına kullanılarak problem üzerinde başarımlar karşılaştırılması yapılmıştır. Buna ilaveten, müşterilerin geri ödeme davranışlarının problem üzerindeki önemini analiz etmek için kullanılan öznetelik vektörü, müşterilerin farklı ürünlerdeki ödeme davranışları ile zenginleştirilerek analizler gerçekleştirilmiştir. Elde edilen sonuçlar, müşterilerin ödeme davranışlarının, müşteri riski analiz probleminde önemli rol oynadığını göstermiştir. Son olarak ise sistemin başarımlarını artırmak için yöntemlerin farklı kombinasyonları ile harmanlanmış sistemler oluşturulmuş ve bu harmanlanmış sistem başarımlarının tek başına kullanılan yöntemlerden daha iyi sonuçlar verdiği gözlemlenmiştir.

#### KAYNAKÇA

- [1] C.-L. Huang, M.-C. Chen, and C.-J. Wang, “Credit scoring with a data mining approach based on support vector machines,” *Expert systems with applications*, vol. 33, no. 4, pp. 847–856, 2007.
- [2] L. Zhou, K. K. Lai, and L. Yu, “Least squares support vector machines ensemble models for credit scoring,” *Expert Systems with Applications*, vol. 37, no. 1, pp. 127–133, 2010.
- [3] G. Derelioğlu and F. Gürgeç, “Knowledge discovery using neural approach for sme’s credit risk analysis problem in turkey,” *Expert Systems with Applications*, vol. 38, no. 8, pp. 9313–9318, 2011.
- [4] L. C. Thomas, J. Ho, and W. T. Scherer, “Time will tell: behavioural scoring and the dynamics of consumer credit assessment,” *IMA Journal of Management Mathematics*, vol. 12, no. 1, pp. 89–103, 2001.
- [5] H. S. Kim and S. Y. Sohn, “Support vector machines for default prediction of smes based on technology credit,” *European Journal of Operational Research*, vol. 201, no. 3, pp. 838–846, 2010.
- [6] J. Banasik, J. N. Crook, and L. C. Thomas, “Not if but when will borrowers default,” *Journal of the Operational Research Society*, vol. 50, no. 12, pp. 1185–1190, 1999.
- [7] H. J. Noh, T. H. Roh, and I. Han, “Prognostic personal credit risk model considering censored information,” *Expert Systems with Applications*, vol. 28, no. 4, pp. 753–762, 2005.
- [8] W. Wang, “How the small and medium-sized enterprises’ owners’ credit features affect the enterprises’ credit default behavior?” *E3 Journal of Business Management and Economics*, vol. 3, no. 2, pp. 090–095, 2012.
- [9] D. H. Wolpert, “Stacked generalization,” *Neural Networks*, vol. 5, pp. 241–259, 1992.
- [10] C. Cortes and V. Vapnik, “Support-vector networks,” *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, Sep. 1995.
- [11] N.-C. Hsieh and L.-P. Hung, “A data driven ensemble classifier for credit scoring analysis,” *Expert systems with Applications*, vol. 37, no. 1, pp. 534–545, 2010.
- [12] G. Wang, J. Hao, J. Ma, and H. Jiang, “A comparative assessment of ensemble learning for credit scoring,” *Expert systems with applications*, vol. 38, no. 1, pp. 223–230, 2011.
- [13] L. Breiman, “Random forests,” *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, Oct. 2001.
- [14] J. H. Friedman, “Stochastic gradient boosting,” *Computational Statistics & Data Analysis*, vol. 38, no. 4, pp. 367–378, 2002.
- [15] N. Sarlija, M. Bensic, and M. Zekic-Susac, “Comparison procedure of predicting the time to default in behavioural scoring,” *Expert Systems with Applications*, vol. 36, no. 5, pp. 8778–8788, 2009.
- [16] M. Jahrer, A. Töschner, and R. Legenstein, “Combining predictions for accurate recommender systems,” in *Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ser. KDD ’10. New York, NY, USA: ACM, 2010, pp. 693–702.