

Kredi Riski Tahmininde Yapay Sinir Ağları ve Lojistik Regresyon Analizi Karşılaştırılması

Comparison of Artificial Neural Networks and Logistic Regression Analysis in the Credit Risk Prediction

Hüseyin BUDAK

Türkiye Vakıflar Bankası T.A.O. Bireysel Kredi
İzleme Müdürlüğü
huseyin.budak@vakifbank.com.tr

Semra ERPOLAT

Yrd. Doç. Dr., Mimar Sinan G.S.Ü. İstatistik Bölümü
serpolat@msgsu.edu.tr

ÖZET

Kredi değerlendirme, finansal sistemdeki sınırlı kaynakların daha verimli kullanılabilmesi açısından, bankalar için oldukça önemli bir konudur. Bankalar, kredi değerlendirmesinde bir çok yöntem kullanmaktadır. Bu yöntemlerden bir tanesi de, kredi talep eden müşterinin kredisini düzenli ödeyip ödemeyeceğinin tahmin edilmesidir. Bu çalışmada, bankaların kredi risklerini öngörmelerine yardımcı olması amacıyla, kredi talep eden müşterilerin ödeme alışkanlıklarının düzenli olup olmayacağının tahmin edilmesi için yapay sinir ağları ve lojistik regresyon analizi kullanılmıştır. Çalışma sonucunda, yapay sinir ağı yönteminin müşterilerin ödeme alışkanlıklarının düzenli olup olmayacağını tahmin etme gücü lojistik regresyon modelinden daha üstün olduğu tespit edilmiştir.

Anahtar sözcükler: Yapay Sinir Ağları, Lojistik Regresyon, Kredi Riski Tahmini

ABSTRACT

Credit scoring is a vital topic for Banks since there is a need to use limited financial sources more effectively. There are several credit scoring methods that are used by Banks. One of them is to estimate whether a credit demanding customer's repayment order will be regular or not. In this study, artificial neural networks and logistic regression analysis have been used to provide a support to the Banks' credit risk prediction and to estimate whether a credit demanding customers' repayment order will be regular or not. The results of the study showed that artificial neural networks method is more reliable than logistic regression analysis while estimating a credit demanding customer's repayment order.

Keywords: Artificial Neural Networks, Logistic Regression, Credit Risk Prediction

1. GİRİŞ

Bankaların karşı karşıya oldukları temel risk faktörlerinden birisi kredi riskidir. Kredi riski; borçlu kişi veya kuruluşun anlaşma şartları dahilinde taahhüt ettiği yükümlülükleri yerine getirememesi olarak tanımlanabilir.

Kredi vermek çoğu bankanın en temel faaliyetidir. Kredi işlemleri, bankaların müşterilerin kredibilitesi konusunda doğru kararlar vermelerini gerektirir. Bu gereklilik bankaları, kredi talep eden müşterilerin kredibilitelerinin tahmin edilmesi için bir çok yöntem kullanmaya zorlamıştır. Önceleri bankalar, müşterilerle bizzat görüşerek müşterilerin kredibiliteleri hakkında elde ettikleri sübjektif kanaatlerine ve uzun yıllar çalışmış bankacıların tecrübelerine dayalı kararlar alırken, bilgisayar teknolojisinin ve istatistiksel tekniklerin hızla gelişmesiyle kredibilite değerlendirmesinde istatistiksel yöntemler kullanmaya başlamışlardır. Bu yöntemlerden biri de, kredi talep eden müşterilerinin ödeme alışkanlıklarını düzenli olup olmayacağını tahmin edilmesidir.

Bu çalışmada, kredi talep eden müşterilerinin ödeme alışkanlıklarını düzenli olup olmayacağını tahmin edilmesinde, yapay sinir ağları ve lojistik regresyon analizinin tahmin etme güçlerinin tespit edilmesi amaçlanmıştır.

2. YAPAY SİNİR AĞLARI

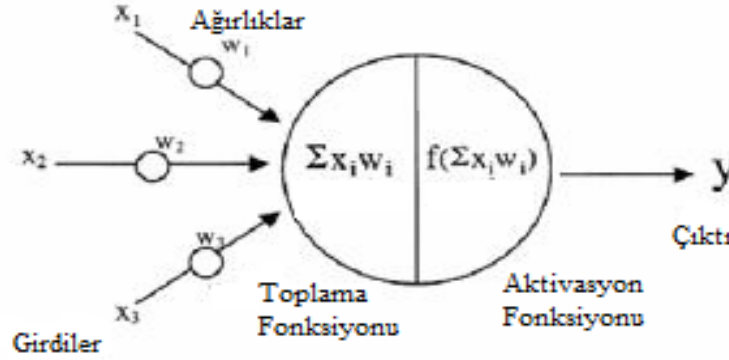
Yapay sinir ağları (YSA), insan beyninden esinlenerek geliştirilmiş, ağırlıklı bağlantılar aracılığıyla birbirine bağlanan ve her biri kendi belleğine sahip işlem elemanlarından oluşan paralel ve dağıtılmış bilgi işleme yapılarıdır. Yapay sinir ağları, başka bir deyişle, biyolojik sinir ağlarını taklit eden bilgisayar programlarıdır (Elmas, 2010). Bu bilgisayar programları ile insan beyninin fonksiyonları arasında benzerlikler bulunmaktadır. Bu yüzden YSA sistemine “insan beyninin modeli” denilebilir.

İnsan beynine benzer şekilde öğrenme, ilişkilendirme, sınıflandırma, genelleme, özellik belirleme ve optimizasyon gibi konularda uygulanmaktadır. Sinirler, örneklerden elde ettikleri bilgiler ile kendi deneyimlerini oluşturur ve daha sonra benzer kararları verirler (Öztemel, 2003).

2.1. Yapay Sinir Hücresi

Yapay sinir ağları, biyolojik sinir hücrelerinin basit bir modeli olan, yapay sinir hücrelerinden meydana gelmektedir. Yapay sinir hücreleri ayrıca düğüm, birim veya işlemci elemanı olarak da adlandırılmaktadır. Yapay sinir hücrelerin de, biyolojik sinir hücrelerine benzer şekilde, girdi sinyallerini aldıkları, bu sinyalleri toplayıp işledikleri ve çıktıları

ilettikleri bölümleri bulunmaktadır. Yapay sinir hücreleri, biyolojik sinir sisteminin doğal yapısının matematiksel olarak modellenmesiyle oluşmaktadır.



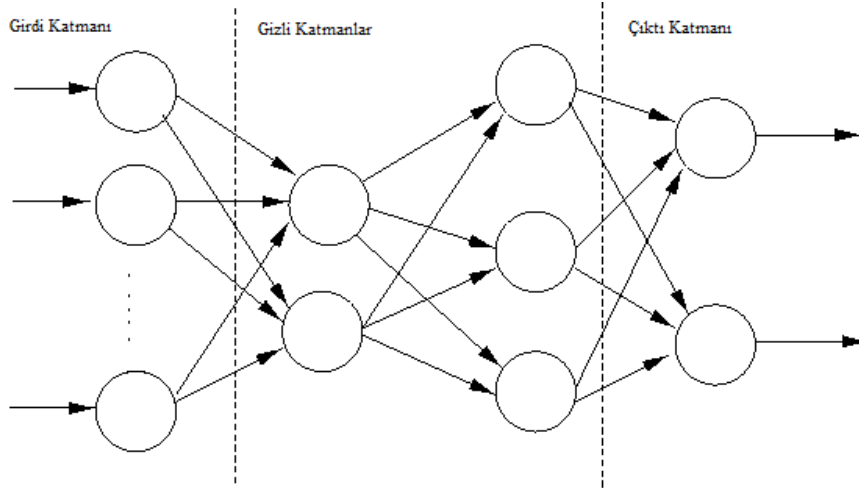
Şekil 1. Yapay sinir hücresi yapısı (Ion, Munteanu & Cocina, 2009)

Girdiler dış ortamdan aldıkları bilgiyi sinire getirirler. Daha sonra bu girdiler, yapay sinir tarafından alınan girdilerin sinir üzerindeki etkisini belirleyen katsayılar olan, ağırlıklar ile çarpılır ve bu çarpımların toplamaları toplama fonksiyonu aracılığıyla eşik değeri (θ_i) ile toplanır. Toplama fonksiyonunun sonucu da aktivasyon (etkinlik) fonksiyonundan geçirilerek çıktı elde edilir (Elmas, 2010).

2.2. Yapay Sinir Ağlarının Yapısı

Yapay sinir ağları, yapay sinir hücrelerinin birbirleri ile çeşitli şekillerde bağlanmalarından oluşur. Hücre çıktıları, ağırlıklar üzerinden diğer hücrelere ya da kendisine girdi olarak bağlanabilir. Hücrelerin bağlantı şekillerine, öğrenme kuralına ve aktivasyon fonksiyonlarına göre çeşitli YSA modelleri geliştirilmiştir.

En basit yapay sinir ağı modelinde tek katman ve tek sinir bulunmaktadır. Bu yapay sinir ağlarının birden çok girdi ve tek bir çıktı bulunmaktadır. Çıktı değeri 1 veya 0 olmalıdır. Bu tip modeller genellikle nesneleri iki ayrı sınıfa ayırmak için kullanılmaktadır. Ancak gizli katmanı olmayan sadece girdi ve çıktı katmanı olan ağlar, karmaşık işlevleri hesaplama yeteneğinden yoksundur. Bu nedenle karmaşık hesaplamalar için oluşturulan ağlarda en az bir gizli katman bulunmalıdır (Elmas, 2010). Bu nedenle, bir yapay sinir ağında, birbirleriyle bağımlı sinirlerin yer aldığı girdi katmanı, gizli katman ve çıktı katmanı olmak üzere üç temel katmandan bulunmaktadır.



Şekil 2. Örnek ileri beslemeli yapay sinir ağı yapısı

Yapay sinir ağının ilk katmanı olan girdi katmanı, dışarıdan gelen verilerin yapay sinir ağına alınmasını sağlar. Bu veriler, istatistikte bağımsız değişkene karşılık gelen girdi değişkenlerinden oluşur. Son katman çıktı katmanı olarak adlandırılır ve istatistikte bağımlı değişkenlere karşılık gelen çıktı değişkenlerinden oluşur. Modeldeki diğer katmanlar ise girdi katmanı ve çıktı katmanı arasında yer alır ve gizli katmanlar olarak adlandırılır. Gizli katmanların dış ortamla bağlantıları yoktur. Yalnızca girdi katmanından gelen sinyalleri alırlar ve çıktı katmanına sinyal gönderirler.

Ağ üzerindeki her birim alt düzeyden gelen girdi değerlerini alır ve ağırlıklı toplamları hesaplar. Girdi verileri bağlantı ağırlıkları ile çarpılarak ağ üzerinde ilerler. Ağırlıklar başlangıçta rastgele olarak belirlenir, böylece sistemin başlangıç noktasında kilitlenmemesi sağlanmış olur (Baxter, 2001). Çarpım sonuçları ilk gizli katmanda toplanır. Sonuç aktivasyon fonksiyonundan (genellikle, 0 ile 1 arasında değişen çıktı sonuçları veren, sigmoid fonksiyonu tercih edilir) geçirilir. Sinyaller çıktı katmanına da gizli katmanlarda olduğu gibi iletilerek süreç tamamlanır (Deboeck, 1994).

3. LOJİSTİK REGRESYON

Regresyon analizi, bağımlı değişken ile bağımsız değişkenler arasındaki ilişkinin fonksiyonel seklini belirleyerek tahmin etmek amacıyla kullanılan bir yöntemdir (Gujarati, 1999). Doğrusal regresyon analizinin uygulanabilmesi için bazı varsayımların sağlanması gerekmektedir. Bu varsayımlardan bir tanesi bağımlı değişkenin sürekli olmasıdır. Ancak, sosyal bilimlerle ilgili uygulamalarda bağımlı değişken genellikle kategorik ve iki düzeylidir. Bağımlı değişkenin iki düzeyli olduğu durumlarda lojistik regresyon analizi kullanılmaktadır.

k tane bağımsız değişken için doğrusal regresyon modeli (1) no.lu denklemdeki gibi yazılır.

$$y_i = \beta_0 + \beta_1 \cdot X_1 + \beta_2 \cdot X_2 + \dots + \beta_k \cdot X_k + \epsilon \quad (1)$$

Bağımlı değişkenin alabileceği değerlerin 0-1 arasında olmasını sağlamak için bağımsız değişkenler ve bağımlı değişken arasında eğrisel bir ilişki (2) no.lu ifadedeki gibi sağlanır (Gujarati, 1999).

$$E(y_i) = p_i = \frac{\exp(y_i)}{1 + \exp(y_i)} = \frac{1}{1 + \exp(-y_i)} \quad (2)$$

(2) no.lu denklem doğrusal olmadığı için β parametrelerini tahmin etmek zordur. Bu denklem doğrusallaştırıldığında aşağıdaki Logit modeli elde edilir (Gujarati, 1999).

$$L_i = \ln\left(\frac{p_i}{1 - p_i}\right) = \beta_0 + \beta_1 \cdot X_1 + \beta_2 \cdot X_2 + \dots + \beta_k \cdot X_k \quad (3)$$

Bu model parametrelerine göre doğrusal olduğundan parametre tahminleri kolaylıkla yapılabilir. Ayrıca bu model sınıflandırma yapmak için de kullanılır. p_i değeri 0.5'den büyük eşit ise ilgili $y_i = 1$, 0.5'den küçük ise ilgili $y_i = 0$ şeklinde sınıflandırma yapılır (Agresti, 1990).

4. UYGULAMA

Günümüzde bireysel kredi kullanımı gelişen tüketici ihtiyaçları doğrultusunda hızla yükselmektedir. Müşterilerin artan hizmet beklentileri, bunun yanı sıra bankaların risklerini kontrol altında tutmak istemeleri ve bankacılık sektöründe yaşanan iç ve dış rekabet ortamı, yaşanan ekonomik krizler ve olumsuz şartlar nedeniyle son yıllarda devlet tarafından yapılan denetimlerin artması kredi kullandıran kuruluşların "kredi riski" kavramına her geçen gün daha fazla önem vermelerini sağlamıştır.

Bankalar, kredi risklerini minimize etmek için kredilerini düzenli ödeyecek müşterilere kredi vermek isterler. Bu durum, kredi kullandırma aşamasında, kredi talep eden müşterinin kredisini düzenli ödeyip ödemeyeceğinin tespitini gerektirmektedir. Bir kredinin düzenli ödenip ödenmeyeceğinin tespitinde öncelikle kredi geri ödemelerinde yaşanan düzensizliklerin nedenlerini tanımlamak gerekmektedir. Belirlenen nedenler ile yapılabilecek birçok istatistiksel analiz yöntemiyle kredilerinin düzenli ödenip ödenmeyeceğini tahmin etmek mümkündür. Uygulamada sıklıkla tercih edilen yöntem,

kredi müşterilerini "kredisini düzenli ödeyen" ve "kredisini düzenli ödemeyen" gibi iki sınıfa ayıran lojistik regresyon analizidir.

Bu çalışmada, bireysel kredi müşterilerini kredi geri ödeme durumlarına göre sınıflandırmak amacıyla lojistik regresyon analizi ve yapay sinir ağları yöntemi kullanılacaktır.

4.1. Veri Seti ve Değişkenlerin Belirlenmesi

Türk bankacılık sektöründe faaliyet gösteren bir bankadan yaklaşık bir yıl önce bireysel kredi kullanmış olan 1639 kişi rastgele olarak çalışmanın örneklemine seçilmiştir. Seçilen bu müşterilerin son bir yıl içerisindeki kredi geri ödemeleri incelenmesi sonucunda müşteriler "kredisini düzenli ödeyen" ve "kredisini düzenli ödemeyen" gibi iki sınıfa ayrılarak çalışmanın bağımlı değişkeni elde edilmiştir. Daha sonra, uzman görüşleri alınarak, kredi geri ödemelerini etkileyeceği düşünülen aşağıdaki bağımsız değişkenler tespit edilmiştir.

Bağımlı değişken;

$Y = \{1, \text{kredi ödemelerinde son bir yılın içerisinde ard arda 30 günden fazla gecikmesi olan}$
 $0, \text{kredi ödemelerinde son bir yılın içerisinde ard arda 30 günden fazla gecikmesi olmayan}\}$

Bağımsız değişkenler;

X_1 : Kredi tutarı

X_2 : Kredinin vadesi

X_3 : Aylık geliri

X_4 : Kredi tutarı / aylık geliri

X_5 : Kefalet (1: kredide kefil var, 0: kredide kefil yok)

X_6 : İpotek (1: kredide ipotek alınmış, 0: kredide ipotek alınmamış)

X_7 : Meslek

X_8 : Yaş

X_9 : Medeni durum (1: Evli, 2: Bekar, 3: Dul)

Tablo 1. Meslek grupları

Kod	Meslek	Kod	Meslek	Kod	Meslek
1	Memur (kamu)	7	Esnaf	13	Mali müşavir / muhasebeci
2	İşçi	8	Mimar / mühendis	14	Teknik eleman
3	Memur (özel)	9	Hukukçu (avukat, hakim, savcı)	15	Öğretim görevlisi
4	Emniyet görevlisi	10	Ev hanımı	16	Yönetici
5	Emekli	11	Bankacı	17	Çiftçi
6	Silahlı kuvvetler mensubu	12	Tıp doktoru	18	Diğer

4.2. Analiz

Yapay sinir ağı uygulaması için çok katmanlı yapay sinir ağı modeli kullanılmıştır. Yapay sinir ağı modeli girdi katmanı, gizli katman ve çıktı katmanı olmak üzere üç tabakadan oluşmaktadır. Analiz için veri setinde yer alan veriler rastgele olarak eğitim ve test seti olarak iki gruba ayrılmıştır. Eğitim seti verinin % 69,6'sını ve test seti ise verinin % 30,4'ünü oluşturmaktadır. Verilerin gruplara ayrılması ve analiz işlemleri IBM SPSS Statistics 19 paket programı yardımıyla yapılmıştır.

Tablo 2. Yapay sinir ağları analizi sınıflandırma sonuçları

Gözlenen		Tahmin edilen		
		0	1	Doğruluk yüzdesi
Eğitim	0	601	87	87,4
	1	300	151	33,5
	%	79,1	20,9	66,0
Test	0	284	32	89,9
	1	116	66	36,3
	%	80,3	19,7	70,3

Lojistik regresyon analizi ile müşterilerin kredilerini düzenli ödeme olasılıkları hesaplanmaktadır. Hesaplanan olasılık değeri 0,5'den büyük olan müşteriler "kredisini düzenli ödemeyen" ve olasılık değeri 0,5'den küçük olan müşteriler "kredisini düzenli ödeyen" sınıfına atanmaktadır. Gözlemlerin sınıflara ataması yapıldıktan sonra gerçek durumla karşılaştırılarak doğru sınıflandırma oranları hesaplanmaktadır. Söz konusu analiz işlemleri IBM SPSS Statistics 19 paket programı yardımıyla yapılmıştır.

Tablo 3. Lojistik regresyon analizi sınıflandırma sonuçları

Gözlenen		Tahmin edilen			
		0	1	Toplam	Doğruluk yüzdesi
Y	0	932	73	1005	92,7
	1	499	135	634	21,3
	Toplam	1431	208	1639	65,1

5. SONUÇ

Bankaların, kredi talep eden müşterilere ilişkin kredi geri ödemelerinde gecikme yaşanıp yaşanmayacağını tahmin etmek için kullanılan birçok istatistiksel yöntem

bulunmaktadır. Bu çalışmada kredi riskinin öngörülmesine yönelik kullanılan istatistiksel yöntemlerden biri olan lojistik regresyon analizi ve yapay sinir ağları kullanılarak, yöntemlerin kredi riskini tahmin etme güçleri karşılaştırılmıştır.

Analiz sonuçlarına göre yöntemlerin sınıflandırma başarıları incelendiğinde yapay sinir ağları modelinin doğru sınıflandırma oranının % 70,3 ve lojistik regresyon modelinin doğru sınıflandırma oranının % 65,1 olduğu görülmektedir. Yapay sinir ağları yönteminin kredi riskinin öngörülmesinde lojistik regresyon analizine göre daha üstün olduğu tespit edilmiştir.

Sonuç olarak, müşterilerin kredi geri ödemelerindeki düzeni etkileyen değişkenler bilindiğinde, kredi talep eden yeni müşterilerin gelecekte kredilerini düzenli ödeyip ödemeyeceklerini tahmin etmek için yapay sinir ağları yöntemi, lojistik regresyon analizine alternatif bir yöntem olarak kullanılabilir. Böylelikle kredi talep eden müşterilerin gelecekte kredilerini düzenli ödeyip ödemeyecekleri tahmin edilerek, kredilerini düzenli ödemeyeceği tahmin edilen müşterilere kredi kullandırılmayabilir ya da bu sınıfta yer alan müşterilerin kredileri diğer kredilerden farklı yöntemlerle ve daha yakından izlenebilir. Bu sayede, bankanın üstlenmiş olduğu kredi riski azaltılabilir.

KAYNAKLAR

- Agresti, A. (1990). *Analysis Of Ordinal Categorical Data*, New York: John Wiley & Sons.
- Baxter, C.W. (2001). *Modelling Heuristics from Literature*, CIV E 729 Course Notes, Dept. of Civil and Environmental Engineering, Edmonton: University of Alberta.
- Deboeck, G. J. (1994). *Trading on the Edge*, New York: John Wiley & Sons.
- Elmas, Ç. (2010). *Yapay Zeka Uygulamaları*, Ankara: Seçkin Yayıncılık.
- Gujarati, D.N. (1999). *Temel Ekonometri*, (Şenesen, Ü. & Şenesen, G.G., Çev.), İstanbul: Literatür Yayıncılık. (Orijinal çalışma basım tarihi 1978)
- Güneri, N. & Apaydın, A. (2004). Öğrenci Başarılarının Sınıflandırılmasında Lojistik Regresyon Analizi ve Sinir Ağları Yaklaşımı, *Gazi Üniversitesi Ticaret ve Turizm Eğitim Fakültesi Dergisi*, Yıl 2004, Sayı 1, Sayfa 170-188.
- Ion, R.-M., Munteanu, D. & Cocina, G.-C. (2009). Concept of artificial neural network (ANN) and its application in cerebral aneurism with multi walls carbon nanotubes (MWCNT), *Proceedings of the 10th WSEAS International Conference on Neural Networks*, Page 104-107.
- Öztemel, E. (2003). *Yapay Sinir Ağları*, İstanbul: Papatya Yayıncılık.