

## BANKACILIK SEKTÖRÜNDE BAŞARISIZLIK TAHMİNİNE VERİ MADENCİLİĐİ YAKLAŞIMI

**Peral Toktař, Melek Başak Demirhan**

*Yeditepe Üniversitesi, Sistem Mühendisliđi Bölümü, 34755, İstanbul*

**Özet:** Türkiye’de yařanan ekonomik krizlerin olumsuz etkileri, bankacılık sektörü de dahil olmak üzere birçok sektörde görölmüřtür. Başarısız bankaların sayısında meydana gelen artış, bu bankaların mümkün olduđunca erken saptanabilmesi ve gerekli önlemlerin önceden alınabilmesi için bir erken uyarı sisteminin geliřtirilmesini gerekli kılmıřtır. Bu çalıřma, Türkiye’de faaliyet gösteren bankalar için başarısızlıđı üç yıl öncesinden itibaren tahmin edecek erken uyarı modellerinin geliřtirilmesini amaçlamaktadır. Örneklemede, Türkiye’de faaliyet gösteren ve 34 tanesi 1997-2003 döneminde başarısız olmuř olan 77 ticaret ve kalkınma-yatırım bankası yer almaktadır. Modeller, bankaların mali tablolarından türetilen finansal oranlar kullanılarak kurulmuřtur. Finansal oranlar yapıları geređi birbirleri ile iliřkili olduđu için, ham veriler kullanılmamıřtır. Çoklu bađlantıyı yok etmek için öncelikle faktör analizi uygulanmıř ve elde edilen faktörler kullanılarak modeller kurulmuřtur. Veri madenciliđi, bankaların başarısızlık tahmini çalıřmalarında anlamlı sonuçlar üretebilen çeřitli teknikler sunan ve geliřmekte olan bir alandır. Modeller iki farklı veri madenciliđi tekniđi olan lojistik regresyon analizi ve sinir ađları ile kurulmuřtur. Modellerin geçerlilik testi çapraz-geçerlilik yöntemi kullanılarak gerçekleştirilmiřtir. Sonuçlar, başarısızlık önceden tahmin edildikçe dođru sınıflandırma oranlarının azaldıđını göstermektedir. Bu çalıřmadaki sinir ađı modellerinin tahmin gücü lojistik regresyon modellerine nazaran genellikle daha yüksektir. Sermaye yeterliliđi, aktif kalitesi, likidite, karlılık ve gelir-gider yapısı, bir bankanın ileride başarısız olma olasılıđının sinyallerini veren göstergeler olarak bulunmuřtur.

**Anahtar Kelimeler:** *Banka Başarısızlık Tahmini, Veri Madenciliđi, Sinir Ađları, Lojistik Regresyon, Faktör Analizi*

### A DATA MINING APPROACH TO PREDICT FAILURES IN BANKING SECTOR

**Abstract:** The negative effects of the economic crises in Turkey covered a wide range of sectors, including the banking sector. The increase in the number of bank failures has brought about the need of an early warning system to detect bank failures as early as possible so that necessary precautions can be taken in advance. This study aims at developing early warning models to predict bank failures in Turkey up to three years prior to failure. 77 commercial and development-investment banks in Turkey, 34 of which failed during 1997-2003, take place in the training sample. The models are constructed using the financial ratios derived from the financial statements of the banks. Since financial ratios are inherently correlated with each other, raw data are not used. First factor analysis is applied to avoid multicollinearity and the acquired factors are used to develop the models. Data mining is an evolving area presenting several techniques that can produce significant results in the bank failure prediction studies. The models are developed using two different data mining techniques, namely logistic regression analysis and neural networks. Validation of the models is performed using the leave-one-out cross-validation method. The results indicate that the classification accuracies of the models decrease as the model predicts the failure earlier. The neural network models in this study have generally better predictive abilities than the logistic regression models. The capital adequacy, asset quality, liquidity position, profitability, and income-expenditure structure of a bank at any time are found to be the indicators of its likelihood of failure at a posterior time.

**Keywords:** *Bank Failure Prediction, Data Mining, Neural Networks, Logistic Regression, Factor Analysis*

#### 1. Giriř

Türkiye’de, özellikle Kasım 2000 ve řubat 2001’de yařanan ekonomik krizler, diđer sektörleri olduđu gibi bankacılık sektörünü de önemli ölçüde etkilemiřtir. Krizler sonucunda faiz oranları yükselmiř; bu da, vade uyumsuzluđu olan bankaların fonlama zararlarının artmasına ve portföylerinde bulundurdıkları menkul kıymetlerin deđer kaybetmesine neden olmuřtur. řubat 2001’de dalgalı kur sistemine geçildikten sonra, Türk Lira’sı yabancı paralar karşısında hızla deđer kaybetmeye bařlamıřtır. Bu durum, yüksek açık pozisyonla çalıřan bankaların önemli boyutta kambiyo zararı ile karşı karşıya

kalmasına yol açmıştır. Diğer yandan, ekonomik aktivitelerdeki azalma, takipteki alacakların artmasına neden olmuştur.

Türkiye’de yaşanan ekonomik krizler sonucunda, başarısız bankaların sayısında da bir artış meydana gelmiştir. Bu da, başarısız bankaların mümkün olduğunca erken saptanabilmesi ve gerekli önlemlerin önceden alınabilmesi için bir erken uyarı sisteminin geliştirilmesini beraberinde getirmiştir.

*Veri madenciliği*, büyük miktarda veriden anlamlı bilgilerin çıkarılması olarak tanımlanıp; birçok uygulamada olduğu gibi bankaların başarısızlık tahmini çalışmalarında da anlamlı sonuçlar üretebilen çeşitli teknikler sunan ve gelişmekte olan bir alandır.

## 2. Banka Başarısızlık Tahmin Modelleri

Bu çalışmada, Türkiye’deki bankalar için başarısızlığı üç yıl öncesinden itibaren tahmin edecek erken uyarı modellerinin geliştirilmesi amaçlanmaktadır. Bu nedenle, iki farklı veri madenciliği tekniği olan lojistik regresyon analizi ve sinir ağları kullanılarak her bir teknikle üç ayrı model kurulmuş; başarısızlığı bir yıl öncesinden tahmin eden modeller *bir-yıllık* (1Y), iki yıl öncesinden tahmin eden modeller *iki-yıllık* (2Y) ve üç yıl öncesinden tahmin eden modeller *üç-yıllık* (3Y) olarak adlandırılmıştır.

Türkiye’de faaliyet gösteren ve 34 tanesi 1997-2003 döneminde başarısız olmuş olan 77 ticaret ve kalkınma-yatırım bankası örnekleme dahil edilmiştir. Bu dönem içinde kapanan, Tasarruf Mevduatı Sigorta Fonu’na devredilen, bir başka bankaya devredilen ya da bir başka banka bünyesinde birleştirilen bankalar başarısız olarak alınmıştır. Başarılı bankalar, başarısız banka sayısının en yüksek olduğu 2001 yılından seçilmiştir. 1Y, 2Y ve 3Y modellerde kullanılan örneklem verileri eksikliği yüzünden bazı farklılıklar göstermektedir.

Modeller, bankaların mali tablolarından türetilen finansal oranlar kullanılarak kurulmuş; finansal tablolar, Türkiye Bankalar Birliği’nin yayınladığı yıllık raporlardan alınmıştır. Veriler, 1Y, 2Y ve 3Y modeller için sırasıyla, başarı ya da başarısızlık yılından bir, iki ve üç yıl önce yayınlanan finansal tablolardan elde edilmiştir. Kullanılan bağımsız değişkenler Tablo 1’de verilmiştir.  $X_1$ ,  $X_2$ ,  $X_3$ ,  $X_4$ , *sermaye yeterliliği*;  $X_5$ ,  $X_6$ ,  $X_7$ ,  $X_8$ , *aktif kalitesi*;  $X_9$ ,  $X_{10}$ ,  $X_{11}$ , *likidite*;  $X_{12}$ ,  $X_{13}$ ,  $X_{14}$ , *karlılık*;  $X_{15}$ ,  $X_{16}$ ,  $X_{17}$ ,  $X_{18}$  ve  $X_{19}$  ise *gelir-gider yapısı* ile ilgili göstergeleridir. Bağımlı değişken olarak ise, başarısız bankalar için 0, başarılı bankalar için 1 değerini alan 0/1 değişkeni kullanılmıştır.

Tablo 1. Bağımsız değişken olarak kullanılan finansal oranlar

Değişken	Tanım
$X_1$	(Özkaynak + Toplam kar) / Toplam aktifler
$X_2$	(Özkaynak + Toplam kar) / (Mevduat + Mevduat dışı kaynaklar)
$X_3$	Net çalışma sermayesi / Toplam aktifler
$X_4$	(Özkaynak + Toplam kar) / (Toplam aktifler + Gayrinakdi krediler)
$X_5$	Toplam krediler / Toplam aktifler
$X_6$	Takipteki krediler / Toplam krediler
$X_7$	Duran aktifler / Toplam aktifler
$X_8$	Yabancı para aktifler / Yabancı para pasifler
$X_9$	Likit aktifler / Toplam aktifler
$X_{10}$	Likit aktifler / (Mevduat + Mevduat dışı kaynaklar)
$X_{11}$	Yabancı para likit aktifler / Yabancı para pasifler
$X_{12}$	Net kar (Zarar) / Ortalama toplam aktifler
$X_{13}$	Net kar (Zarar) / Ortalama özkaynaklar
$X_{14}$	Net kar (Zarar) / Ortalama ödenmiş sermaye
$X_{15}$	Takipteki alacak sonrası net faiz geliri / Ortalama toplam aktifler
$X_{16}$	Faiz gelirleri / Faiz giderleri
$X_{17}$	Faiz dışı gelirler / Faiz dışı giderler
$X_{18}$	Toplam gelirler / Toplam giderler
$X_{19}$	Takipteki alacaklar karşılığı / Toplam krediler

Finansal oranlar yapıları gereği birbirleri ile ilişkili olduğu için, ham verinin kullanılması sonuçların yanlış çıkmasına yol açar. Bu nedenle, çoklu bağlantıyı yok etmek için öncelikle *faktör analizi* uygulanmış ve elde edilen faktör puanları kullanılarak modeller kurulmuştur. Faktör analizi sadece çoklu bağlantıyı yok etmek için uygulandığından, faktör sayısı bağımsız değişken sayısına eşit alınmıştır.

Modellerin kurulmasında iki farklı veri madenciliği tekniği olan lojistik regresyon analizi ve sinir ağları kullanılmıştır. *Lojistik regresyon analizi* istatistiksel bir veri madenciliği tekniği olup, metrik

olmayan bir bağımlı değişkenin bağımsız değişkenler kullanılarak tahmin edilmesi esasına dayanır. Normalde bağımlı değişken ile bağımsız değişkenler arasındaki ilişki doğrusal olmamakla birlikte, bağımlı değişkene logit transformasyonu uygulanarak aradaki ilişki doğrusal hale getirilir. *Sinir ağları* (yapay sinir ağları) ise insan beyninin çalışma ilkelerinin bilgisayarlar üzerinde taklit edilmesi fikri ile ortaya çıkmıştır. Nöronlardan oluşan yüksek performanslı paralel dağıtımlı işlemciler olarak tanımlanan sinir ağları, insan beyni ile bazı benzerlikler gösterir. Her ikisi de bilgiyi çevrelerinden bir öğrenme süreci sonucunda edinirler ve edindikleri bilgileri nöronlar arasındaki sinaptik bağlantılar aracılığıyla saklarlar.

Bu çalışmadaki 1Y, 2Y ve 3Y lojistik regresyon modelleri SPSS paket programı kullanılarak kurulmuştur. Modellerin kurulmasında *adimsal regresyon metodu* uygulanmıştır. 1Y, 2Y ve 3Y sinir ağı modelleri ise MATLAB paket programı kullanılarak oluşturulmuştur. Her üç modelde de 19 nörondan oluşan bir giriş katmanı, beşer nörondan oluşan iki gizli katman ve iki nörondan oluşan bir çıkış katmanı bulunmaktadır. Diğer bir deyişle, her üç sinir ağı da 19-5-5-2 yapısındadır. 1Y ve 2Y sinir ağı modelleri *tek-basamaklı sekant algoritması* ile, 3Y sinir ağı modeli ise *Broyden-Fletcher-Goldfarb-Shanno algoritması* kullanılarak eğitilmiştir. Modellerin doğru sınıflandırma oranları Tablo 2’de verilmiştir.

Tablo 2. Modellerin doğru sınıflandırma oranları

	1Y	2Y	3Y
Lojistik regresyon modelleri	%95,8	%93,0	%87,1
Sinir ağı modelleri	%100,0	%100,0	%95,7

Bir modelin geçerliliğinin test edilmesi için en iyi yöntem yeni veri toplamak ya da örnekleme ikiye bölmektir. Ancak verinin az olmasından dolayı, bu çalışmada modellerin geçerlilik testi *çapraz-geçerlilik yöntemi* kullanılarak gerçekleştirilmiştir. Modellerin geçerlilik oranları Tablo 3’de verilmiştir.

Tablo 3. Modellerin geçerlilik oranları

	1Y	2Y	3Y
Lojistik regresyon modelleri	%83,1	%73,2	%71,4
Sinir ağı modelleri	%84,5	%76,1	%70,0

### 3. Sonuçlar

Bu çalışma, başarısız bankaların mümkün olduğunca erken saptanabilmesi ve gerekli önlemlerin önceden alınabilmesi için bir erken uyarı sisteminin geliştirilebileceğini ortaya koymaktadır. Elde edilen sonuçlar, başarısızlık önceden tahmin edildikçe modellerin başarılı ve başarısız bankaları doğru sınıflandırma oranlarının azaldığını, ancak üç yıl öncesinden dahi başarısız bankaların %71,4 gibi yüksek bir oranla doğru tahmin edilebileceğini göstermektedir. Bu çalışmadaki sinir ağı modellerinin tahmin gücü lojistik regresyon modellerine nazaran genellikle daha yüksektir. Sermaye yeterliliği, aktif kalitesi, likidite, karlılık ve gelir-gider yapısı, bir bankanın ileride başarısız olma olasılığının sinyallerini veren göstergeler olarak bulunmuştur.

### Kaynaklar

**Bankacılık Düzenleme ve Denetleme Kurumu**, *Yıllık Rapor*, 2001.

**Battiti, R.**, “First- and Second-Order Methods for Learning: Between Steepest Descent and Newton’s Method”, *Neural Computation*, 4(2): 141-166, 1992.

**Efe, Ö. ve Kaynak, O.**, *Yapay Sinir Ağları ve Uygulamaları*, Boğaziçi Üniversitesi, İstanbul, 2000.

**Han, J. ve Kamber, M.**, *Data Mining: Concepts and Techniques*, Morgan Kaufmann Publishers, Inc., San Francisco, California, 2001.

**Haykin, S.**, *Neural Networks: A Comprehensive Foundation*, Prentice-Hall, Inc., Upper Saddle River, New Jersey, 1999.

**Johnson, R. A. ve Wichern, D. W.**, *Applied Multivariate Statistical Analysis*, Prentice-Hall, Inc., Upper Saddle River, New Jersey, 1998.

**Neter, J. ve diğerleri**, *Applied Linear Statistical Models*, The McGraw-Hill Companies, Inc., Boston, Massachusetts, 1996.