

## Veri Zarflama Analizi Tabanlı Yeni Bir Hibrid İki Gruplu Sınıflandırma Modeli

H. Hasan ÖRKÜ\*<sup>1</sup>, Mustafa İsa DOĞAN

<sup>1</sup>Gazi Üniversitesi, Fen Fakültesi, İstatistik Bölümü, 06500, Teknikokullar, Ankara

(Alınış Tarihi: 08.02.2015, Kabul Tarihi: 13.08.2015)

### Anahtar Kelimeler

Sınıflandırma problemi  
Veri Zarflama Analizi  
BCC modeli  
İki aşamalı yaklaşım

**Özet:** Bu çalışmada iki gruplu sınıflandırma problemlerinin çözümünde kullanılabilecek yeni bir sınıflandırma modeli geliştirilmiştir. Bu model Veri Zarflama Analizi BCC modeline dayanan Pendharkar ve Troutt (2014) modeli ile Sueyoshi (2004) tarafından önerilen iki aşamalı sınıflandırma modelinin bir karmasıdır. Çalışmanın amacı, BCC modelindeki parçalı doğrusal etkinlik sınırı ve iki aşamalı detaylı inceleme fikri sayesinde iki gruplu sınıflandırma problemlerini ele almaktır. Önerilen yeni yaklaşım Pendharkar ve Troutt (2014)'den alınan bir örnek üzerinde ayrıntılı olarak incelenmiş ve ayrıca yapılan simülasyon çalışmasından önerilen yöntemin sınıflandırma performansının diğer iki yöntemden daha iyi olduğu gözlenmiştir.

## A New Hybrid Two-Group Classification Model Based on the Data Envelopment Analysis Model

### Keywords

Classification problem  
Data envelopment analysis  
BCC model  
Two-stage approach

**Abstract:** In this study, a new classification model can be used the two-group classification problem is developed. This approach is mixed of the Pendharkar and Troutt (2014) model bases on the Data Envelopment Analysis BCC model and the two-stage classification model proposed by Sueyoshi (2004). The aim of this study is handled the two-group classification problem with the help of piecewise linear efficiency frontier in the BCC model and two-stage detailed examination idea. The proposed new approach is detailed examined on an example taken by Pendharkar and Troutt (2014). Also, it is observed that the classification performance of proposed method is better than the other two methods from the simulation study.

### 1. Giriş

Bilimin en temel yöntemlerinden birisi olan sınıflandırma, insanoglu tarafından üstlenilen en eski bilimsel uğraşlardan biridir. Sınıflandırma problemi, belirli bir sınıfa ait olduğu bilinen örneklerin kullanılarak, yeni örneklerin sınıflarının belli bir doğruluk ile belirlenmesidir. Sınıflandırma problemleri "ayırma analizi" teknikleri ile incelenmektedir. Ayırma analizi sosyal bilimlerde, finans, pazarlama ve muhasebe gibi iş alanlarında, tıbbi teşhis, hava tahmini, kredi kartı veya kredi almak için yapılan başvuruların geri ödemesinin yapılıp

yapılmayacağı tahmin edildiği kredi değerlendirme, sahtekârlık belirleme gibi bankacılık alanında ve biyoloji gibi taksonomi ve sınıflama analizlerinin içerildiği diğer alanlarda başarı ile uygulanmaktadır.

Bireyleri iki ya da daha fazla gruba sınıflandırmak için farklı kriterler kullanılmaktadır. Fisher'in, iki grup için ortaya atmış olduğu doğrusal ayırma fonksiyonu, ayırma analizinde en çok kullanılan kriterdir. Fisher (1936), çok değişkenli normal dağılım varsayımı altında iki ya da daha fazla gruptan gözlenmiş birimleri gruplardan birine sınıflandırmak için mevcut

\* İlgili yazar: hhorkcu@gazi.edu.tr

değişkenler üzerinden tanımlanacak doğrusal fonksiyonları önermiştir. Öyle ki bu doğrusal fonksiyonların katsayılar vektörü, gruplar arası farklılığı maksimum yapacak biçimde alınır. Fisher (1936), bu problemin çözümü için söz konusu değişkenlerin doğrusal fonksiyonlarından hareket etmiştir. Şöyle ki;

$$Y = \alpha_1 x_1 + \alpha_2 x_2 + \dots + \alpha_k x_k = \underline{\alpha}' \underline{x} \quad (1.1)$$

olsun. İki gruplu doğrusal ayırma analizinde,  $k$  boyutlu  $\alpha$  ayırma vektörü ve  $c$  sabiti belirlenir.  $\underline{\alpha}' \underline{x} \leq c$  ise ilgili birim 1. gruba, diğer durumda 2. gruba ait olarak sınıflandırılır. Bu doğrusal fonksiyon, farklı grupların kovaryans matrislerinin aynı olması durumunda uygundur. Aksi halde karesel fonksiyonların kullanımı analiz sonuçlarının sağlıklı olmasını sağlayacaktır (Lachenburch, 1975). İstatistiksel yöntemler kullanılarak elde edilen sonuçlar olasılıklarla ifade edilir fakat bu yöntemlerin kullanılabilmesi için gerekli olan varsayımların sağlanması çoğunlukla imkânsızdır (Anderson, 1984). Sınıflandırma problemlerinin incelenmesinde istatistiksel yöntemlere alternatif olarak çok sayıda matematiksel programlama yöntemi geliştirilmiştir. Matematiksel programlama yöntemlerini kullanmak için yığınla ilgili herhangi bir varsayımın sağlanmasına gerek yoktur. Doğrusal programlama ile sınıflandırma problemlerinin incelenmesi ilk defa Fred ve Glover (1981a, 1981b) tarafından yapılmıştır. Fred ve Glover, iki gruplu sınıflandırma problemleri için, sapmalar toplamının minimizasyonuna dayanan bir model önermişlerdir.  $k$  değişkenli iki gruplu sınıflandırma problemi göz önüne alınsın.  $Z$ ;  $G_1$  ve  $G_2$  gruplarından alınan  $n$  çaplı bir örneğin  $k$  değişken skorlarını gösteren  $k \times n$ 'lik bir matris olsun. Değişken katsayıları  $\alpha_1, \alpha_2, \dots, \alpha_k$  ile ve  $j$ . birime

ait sınıflandırma skoru ise  $S_j = \sum_{i=1}^k \alpha_i Z_{ij}$  olarak

tanımlanır ( $j = 1, 2, \dots, n$ ). Bir birimin ait olduğu grubun tayin edilmesi o birimin sınıflandırma skorunun değerine bağlıdır. Fred ve Glover (1981a) tarafından önerilen basit MSD (Minimization Sum of Deviations-sapmalar toplamının minimizasyonu) sınıflandırma modeli,

$$\text{Min} \sum_{j \in G_1} S_j^+ + \sum_{j \in G_2} S_j^-$$

Kısıtlar:

$$\sum_{i=1}^k \alpha_i Z_{ij} + S_j^+ \geq c \quad j \in G_1 \quad (1.2)$$

$$\sum_{i=1}^k \alpha_i Z_{ij} - S_j^- \leq c \quad j \in G_2$$

biçiminde tanımlanabilir. Burada  $Z_{ij} \geq 0$ ,  $S_j^+ \geq 0$ ,  $S_j^- \geq 0$   $j = 1, 2, \dots, n$ ,  $\alpha_i$  ( $i = 1, 2, \dots, k$ ) ve  $c$  işaretçe serbest (pozitif veya negatif değerler alabilen) değişkenlerdir. (1.2) ile verilen model doğrusal programlama modelidir ve Simpleks algoritması ile çözülebilir. Bu modelin çözülmesi ile  $\alpha_i$  ve  $c$  değerleri elde edilerek, herhangi bir birimin sınıflandırma skoru elde edilebilir. Bir birimin sınıflandırma skoru ( $S_j$ )  $c$ 'ye eşit ya da büyük ise  $G_1$ 'e, diğer durumda  $G_2$ 'ye sınıflandırılır.  $S_j^+ \geq 0$  olması,  $G_1$  grubundaki  $j$ . birimin yanlış sınıflandırıldığını;  $S_j^- \geq 0$  olması,  $G_2$  grubundaki  $j$ . birimin yanlış sınıflandırıldığını göstermektedir.

Matematiksel programlama yöntemleri dağılımdan bağımsız ve sınıflama amaçları için çok kullanışlıdır. Matematiksel programlama yöntemleri ile çoğunlukla iki gruplu sınıflandırma problemleri incelenmiştir (Fred ve Glover (1981a, 1981b, 1986a, 1986b), Markowski ve Markowski (1985), Joachimsthaler ve Stam (1988), Koehler (1989), Koehler ve Erenguc (1990), Glover (1990), Rubin (1990), Lee ve Ord (1990), Stam ve Jones (1990), Stam ve Ragsdale (1992), Hosseini ve Armacost (1994), Lam vd. (1996), Lam ve Moy (1996, 1997, 2002), Sueyoshi (1999, 2001, 2004), Bal vd. (2006a, 2006b), Bal ve Örkücü (2005, 2007, 2011), Örkücü ve Bal (2011)).

Karar verme birimlerinin performanslarının belirlenmesi amacıyla geliştirilmiş olan ve matematiksel programlamaya dayanan Veri Zarflama Analizi de son yıllarda sınıflandırma problemlerinin çözümünde etkin bir şekilde kullanılmaktadır. Sueyoshi (1999), Veri Zarflama Analizi (Data Envelopment Analysis-DEA) ile ayırma analizi arasındaki farklılıkları ve benzerlikleri hedef programlama ışığında incelediği çalışmasında, VZA modellerinden toplamsal VZA modelinin metodolojik gücünü ayırma analizi ile birleştirerek DEA-DA (Data Envelopment Analysis-Discriminant Analysis) adını verdiği yeni bir sınıflandırma modeli önermiştir. Sueyoshi'nin modeli ilk aşamada birimlerin ön bir incelemeye tabi tutulduğu (kesin olarak grup 1'e ait olan, kesin olarak grup 2'ye ait olan ve grubu tam olarak belirlenemeyen yani çakışma durumunda olan birimlerin tespiti) ve ikinci aşamada ise çakışma durumunda olan birimlerin ayrıca değerlendirilerek nihai sınıflandırmanın yapıldığı iki aşamalı bir sınıflandırma modelidir. Sueyoshi daha sonra bu modeli iki gruplu durumda karma tamsayılı modele, çok gruplu durumda ise grupların büyük-orta-küçük şeklinde sıralamasının mümkün olduğu çok özel bir durumda ise hedef programlama ve karma tamsayılı programlama modeline genişletmişlerdir (Sueyoshi; 2001, 2004). Sueyoshi'nin (1999, 2001, 2004) çalışmalarıyla popüler hale gelen iki aşamalı VZA

tabanlı sınıflandırma modelleri iki gruplu sınıflandırma problemlerinde başarıyla kullanılmaktadır. Pendharkar (2011), negatif değerler içeren verileri Radyal Tabanlı Sinir Ağları (RBFN) ile dönüştürdükten sonra VZA'nın BCC modeline dayanan RBFN-DEA ismini verdikleri sınıflandırma modelini önermişlerdir. Pendharkar ve Troutt (2014) bu yaklaşımı pozitif değer içeren veri setlerinde interaktif bir yaklaşımla sunmuşlardır. Bu yaklaşımlarda sınıflandırma modeli çözüldükten sonra birimlerin sınıflandırılmasında Öklid kuralları ve interaktif adını verdikleri bir yaklaşım kullanılsa da birimlerin gruplara tayininde açık kurallar bulunmamaktadır.

Bu çalışmada, Sueyoshi (1999, 2001 ve 2004) modellerindeki iki aşamalı inceleme fikri, Pendharkar ve Troutt (2014) parçalı doğrusal VZA sınıflandırma modeli ile birleştirilerek yeni bir iki gruplu sınıflandırma modeli elde edilmektedir. Önerilen yöntemin birinci aşamasında her iki grup için ayrı ayrı Pendharkar ve Troutt VZA-BCC modeline dayanan parçalı doğrusal etkinlik sınırları elde edilerek birimlerin ön incelemesi tamamlanmaktadır. Yani iki gruplu problemde, 1. grup için oluşturulan parçalı etkinlik sınırının altında kalan ve 2. grup için oluşturulan parçalı etkinlik sınırının üstünde kalan birimler sırasıyla 1. ve 2. gruba doğru sınıflandırılan birimler olacaktır. Yöntemin ikinci aşamasında ise ilk aşamada iki gruptan birine atanamayan birimler Sueyoshi modellerinde önerilen yaklaşımla yeniden ele alınmaktadır.

Bu çalışmanın amacı, Pendharkar ve Troutt (2014) tarafından önerilen VZA-BCC tabanlı iki gruplu sınıflandırma modelini Sueyoshi (2004) tarafından önerilen iki aşamalı inceleme yaklaşımıyla kombine ederek etkin yeni bir iki gruplu sınıflandırma modeli elde etmektir. Bu çalışmanın Pendharkar ve Troutt (2014) ve Sueyoshi (2004) modellerinden temel farkı, her iki yöntemin de güçlü yönlerini kombine ederek VZA'nın BCC modeli sayesinde doğrusal olarak ayıramayan yapıda olan veri setlerinde parçalı doğrusal ayırma sınırları elde edebilmesi ve iki aşamalı yaklaşım fikri ile birimlerin sınıflandırılmasında detaylı bir inceleme yapabilmesidir.

Çalışmanın 2. bölümünde Veri Zarflama Analizi (VZA) ve temel modeller olan CCR ve BCC modelleri genel hatları ile verilmekte, 3. bölümde bu çalışmanın da temelini oluşturan VZA tabanlı Sueyoshi (2004) ve Pendharkar ve Troutt (2014) sınıflandırma modellerine yer verilmektedir. 4. bölümde ise önerilen model yer almakta ve Pendharkar ve Troutt (2014)'den alınan bir örnek üzerinden önerilen modelin işleyişi ise 5. bölümde sunulmaktadır. Çalışmanın 6. bölümünde önerilen yöntemin sınıflandırma performansını literatürde önerilen birkaç yöntemle kıyaslamak için bir 1000 tekrarlı simülasyon çalışmasının sonuçlarına yer verilmiştir. Son bölümde ise sonuçlar kısmına yer verilmiştir.

## 2. Veri Zarflama Analizi

Veri Zarflama Analizi, karar verme birimlerini görelilik olarak bir bütün içinde karşılaştırma yaparak değerlendirip, etkin olan veya olmayan karar verme birimi şeklinde iki sınıfa ayıran, etkin olmayan karar birimi için de etkinsizliğin kaynaklarını ve miktarını belirleyerek karar vericilere yardımcı olan yaklaşımlar bütünüdür. Ortaya çıktığı ilk günden bu güne çok çeşitli alanlarda yaygın kullanım alanı bulmuş ve birçok teknikle de etkileşim halinde gelişmesini devam ettirmektedir.

Veri Zarflama Analizi metodu girdiye ve çıktıya yönelik olarak iki yönlü kullanılabilme özelliğine sahiptir. Girdiye yönelik Etkinlik Analizi modelleri; belirli bir çıktı bileşimini en etkin bir şekilde üretebilmek amacıyla kullanılacak en uygun girdi bileşiminin nasıl olması gerektiğini araştırır. Çıktıya yönelik Etkinlik Analizi modelleri ise belirli bir girdi bileşimi ile en fazla ne kadar çıktı bileşimi elde edilebileceğini araştırır.

Veri Zarflama Analizinde tüm modellerin ortak yanı, hangi KVB'lerin etkinlik sınırını oluşturduklarını, böylece etkinlik sınırının oluşturulmasıyla, etkin ve etkin olmayan KVB'lerin tespit edilmesidir. Modeller arasındaki fark, kullanılan modele göre bu yüzeyin geometrisinde ortaya çıkmaktadır. Sınırın oluşturulmasıyla birlikte sınırın altında kalan etkin olmayan KVB'ler için kullanılmayan kaynaklar tanımlanabilir.

En çok kullanılan Veri Zarflama Analizi modeli Charnes Cooper Rhodes (CCR) ve Banker Charnes Cooper (BCC) modelleridir.

### 2.1. Charnes Cooper Rhodes (CCR) modeli

Veri Zarflama Analizinin ilk önerilen modeli olan CCR modeli, Charnes, Cooper ve Rhodes (1978) tarafından önerilmiştir. Bu model ile ölçeğe göre sabit getiri varsayımı altında teknik etkinlik ölçülmektedir. Bu tip Etkinlik Analiz modellerinde etkinlik sınırı orijinden başlayıp, etkin olan karar verme birim(ler)inden geçen bir doğru ile temsil edilmektedir.

Girdileri çıktılara dönüştüren  $n$  tane karar verme birimi olsun. Her karar verme birimi için girdi ve çıktı çoklukları değişmekle birlikte, kullanılan girdi ve üretilen çıktı miktarları aynıdır. Matematiksel gösterimle  $j$ . karar verme birimi  $s$  boyutlu çıktı vektörü  $y_{rj}$  ( $r = 1, 2, \dots, s$ ) üretmek üzere  $m$  boyutlu girdi vektörü  $x_{mi}$  ( $i = 1, 2, \dots, m$ ) kullanır.

Girdi ve çıktı değişkenleriyle ilişkilendirilen etkinlik ölçüsü (2.1)'de verildiği gibi tanımlanabilir (Cooper vd., 2000).

$$\text{Etkinlik} = \frac{\text{Çıktıların ağırlıklı toplamı}}{\text{Girdilerin ağırlıklı toplamı}} \quad (2.1)$$

Değerlendirilecek karar verme birimini  $o$  indisi ile, diğerlerini ise  $j$  indisi ile gösterilsin. Etkinlik skorları oran formunda,

$$\max h_o = \frac{\sum_{r=1}^s u_r y_{ro}}{\sum_{i=1}^m v_i x_{io}}$$

Kısıtlar:

$$\frac{\sum_{r=1}^s u_r y_{rj}}{\sum_{i=1}^m v_i x_{ij}} \leq 1 \quad (2.2)$$

$$v_i \geq 0 \quad i = 1, 2, \dots, m$$

$$u_r \geq 0 \quad r = 1, 2, \dots, s$$

biçiminde tanımlanabilir.

Burada  $v_i$  ve  $u_r$  sırasıyla girdi ve çıktı ağırlıklarını göstermektedir.  $\sum u_r y_{rj}$  çıktı toplamını,  $\sum v_i x_{ij}$  girdi toplamını göstermektedir. Çıktı/Girdi oranı  $h_o$ , optimal girdi-çıktı ağırlıklarını seçerek maksimum yapılacak amaç fonksiyonudur. Eşitsizlik kısıtı aynı ağırlıklarla tüm karar verme birimlerinin etkinlik oranlarının birim büyüklükten fazla olmasını garanti eder. Çözüm sonunda elde edilen etkinlik dereceleri  $h_o = 1$  ise o-uncu karar verme birimi  $KVB_o$  tam etkindir.

(2.2) ile verilen model yaklaşımı anlatım bakımından uygun olmakla birlikte hesaplama açısından zorluklar içerir. Bu nedenle, daha uygun bir yapıya dönüştürmek için Girdi Yönlü CCR Primal modeli

$$\max w_o = \sum_{r=1}^s u_r y_{ro}$$

Kısıtlar:

$$\sum_{i=1}^m v_i x_{io} = 1$$

$$\sum_{r=1}^s u_r y_{rj} - \sum_{i=1}^m v_i x_{ij} \leq 0 \quad ; \quad j = 1, 2, \dots, n \quad (2.3)$$

$$v_i \geq 0 \quad ; \quad i = 1, 2, \dots, m$$

$$u_r \geq 0 \quad ; \quad r = 1, 2, \dots, s$$

biçiminde önerilmiştir.

Böylece geleneksel doğrusal programlama modeli elde edilmiş ve hesaplama avantajları doğmuştur. (2.3) modeli, bilgisayar yazılımları ile kolayca çözümlenecek durumdadır. (2.3)' de gösterilen modelden de görüleceği üzere, girdi yönlü CCR modelinde sırasıyla her bir KVB'nin çıktıların ağırlıklı ortalaması maksimum yapılmaya çalışılır. Kısıtlarda ise ilgilenilen KVB'nin girdilerinin ağırlıklı ortalaması 1'e eşitlenmiştir, böylece girdilerin ağırlıklı ortalaması her bir KVB için 1 olmaktadır. Daha sonraki kısıt çıktıların ağırlıklı ortalamasının girdilerin ağırlıklı ortalamasından küçük olmasını sağlamaktadır. Bu sayede Çıktı/Girdi oranı her bir karar verici için en fazla 1 olabilir. Bu bilgilerin ışığında bir karar verici için bulunabilecek optimum çıktı ortalaması en fazla 1 olabilir ve bu ise karar vericinin etkin olduğu anlamına gelir. Etkin olmayan, yani etkinlik sınırının altında kalan KVB'ler için çıktıların ağırlıklı ortalaması, yani etkinlik değeri 1'den küçük olacaktır.

$$\text{Etkinlik için referans noktaları} \quad \sum u_r y_{rj} / \sum v_i x_{ij} = 1$$

olan karar birimleridir.  $KVB_o$ 'ya atanan performans katsayısı  $w_o$ , tüm diğer karar vericilerin performansları üzerinden hesaplanır ve  $v_i^*$ ,  $u_r^*$  ağırlıkları bu değeri maksimum yapan ağırlıklardır.  $w_o^* = 1$  olduğunda  $KVB_o$  diğer karar verme birimlerine göre tam etkin sayılır.

## 2.2. Banker Charnes Cooper (BCC) modeli

Banker, Charnes ve Cooper (1984) tarafından önerilmiştir. BCC modeli ölçeğe göre değişen getiri varsayımı altında karar verme birimlerinin etkinliğini ölçmektedir. Banker, Farrell tarafından tanımlanan ve CCR modeli ile bulunan teknik etkinliğin, ölçek etkinliği ile karışmış olduğunu belirlemiş, teknik etkinliğin ölçek etkinliği ve saf teknik etkinlik olarak ayrılması gerektiğini göstermiştir. Bu nedenle ölçeğe göre değişen getiri varsayımı altında BCC modeli ile saf teknik etkinlik bulunabilmektedir.

CCR modelinde orijinden geçen etkinlik doğrusu BCC modelinde orijinden geçmek zorunda değildir. Bu yapıyla BCC modeli CCR modelinden ayrılmaktadır. Modellerin diğer değişkenler açısından yorumunda bir farklılık yoktur. BCC modeli (2.4) ile verilmektedir.

$$\min z_o = \theta$$

Kısıtlar:

$$\sum_{j=1}^n \lambda_j y_{rj} - s_r^+ = y_{ro} \quad ; \quad r = 1, 2, \dots, s$$

$$\theta x_{io} - \sum_{j=1}^n \lambda_j x_{ij} - s_i^- = 0 \quad ; \quad i = 1, 2, \dots, m \quad (2.4)$$

$$\sum_{j=1}^n \lambda_j = 1$$

$$\lambda, s^+, s^- \geq 0$$

### 3. VZA Tabanlı Sınıflandırma Modelleri: Sueyoshi (2004) ve Pendharkar ve Troutt (2014) Modelleri

Bu bölümde, önerilen modelin dayandığı Sueyoshi (2004) ve Pendharkar ve Troutt (2014) modellerine yer verilmektedir.

Sueyoshi (2004) tarafından önerilen DEA-DA modeli çakışma durumunun belirlendiği ve sınıflandırmanın yapıldığı iki aşamadan oluşmaktadır.

*Aşama 1: Çakışmayı belirleme ve sınıflandırma*

$$\text{Min} \sum_{j \in G_1} S_{1j}^+ + \sum_{j \in G_2} S_{2j}^-$$

Kısıtlar:

$$\sum_{i=1}^k (\lambda_i^+ - \lambda_i^-) Z_{ij} + S_{1j}^+ - S_{1j}^- = d + 1 \quad j \in G_1$$

$$\sum_{i=1}^k (\lambda_i^+ - \lambda_i^-) Z_{ij} + S_{2j}^+ - S_{2j}^- = d \quad j \in G_2 \quad (3.1)$$

$$\sum_{i=1}^k (\lambda_i^+ + \lambda_i^-) = 1$$

$$S_{1j}^+, S_{1j}^-, S_{2j}^+, S_{2j}^-, \lambda_i^+, \lambda_i^- \geq 0, d \text{ işaretçe serbest.}$$

formüle edilebilir. (3.1) modelinde  $S_{1j}^+$  ve  $S_{1j}^-$  sırasıyla  $G_1$  grubuna ilişkin,  $d$  ayırma skorundan

$\sum_{i=1}^k \alpha_i Z_{ij}$  parçalı doğrusal ayırma fonksiyonuna olan pozitif ve negatif sapmaları göstermektedir. Benzer olarak  $S_{2j}^+$  ve  $S_{2j}^-$ 'de sırasıyla  $G_2$  grubuna ilişkin,  $d$

ayırma skorundan  $\sum_{i=1}^k \alpha_i Z_{ij}$  parçalı doğrusal ayırma

fonksiyonuna olan pozitif ve negatif sapmaları göstermektedir.  $S_{1j}^+ > 0 (j \in G_1)$  ve  $S_{2j}^- > 0 (j \in G_2)$  yanlış sınıflandırmanın varlığını

göstermektedir.  $\lambda_i^* (= \lambda_i^+ - \lambda_i^-)$  ve  $d^*$ , 3.2

modelinin optimum çözümleri olsunlar. Değeri  $Z_{im}$  ile gösterilmiş  $m$ . birimin  $G_1, G_2$  ya da  $G_1 \cap G_2$  (çakışma) durumlarından hangisine ait olduğu aşağıda verilen kriterle belirlenir.

$\sum_{i=1}^k \lambda_i^* Z_{im} > d^* + 1$  ise  $m$ . birimin  $G_1$  grubuna ait olduğu sonucuna varılır.

$d^* + 1 > \sum_{i=1}^k \lambda_i^* Z_{im} > d^*$  ise  $m$ . birimin  $G_1 \cap G_2$ 'ye ait olduğu sonucuna varılır.

$d^* \geq \sum_{i=1}^k \lambda_i^* Z_{im}$  ise  $m$ . birimin  $G_2$  grubuna ait olduğu sonucuna varılır.

Yukarıdaki kriterlere göre, tüm birimler ( $G_1 \cup G_2$ ) aşağıda verilen alt kümelerle ayrılabilir.

$$C_1 = \left\{ j \in G_1 \mid \sum_{i=1}^k \lambda_i^* Z_{ij} > d^* + 1 \right\}$$

$$C_2 = \left\{ j \in G_2 \mid \sum_{i=1}^k \lambda_i^* Z_{ij} \leq d^* \right\}$$

$$D_1 = G_1 - C_1 \text{ ve } D_2 = G_2 - C_2$$

*Aşama 2: Çakışmayı ele alma*

Çakışmanın ( $G_1 \cap G_2$ ) varlığı tanımlandıktan sonra, bir yeni ayırma skoru ( $c$ ) modelin içine sadece  $D_1 \cup D_2$  alt grubunda yer alan birimleri sınıflandıracak şekilde konulur.

$$\text{Min} \sum_{j \in D_1} S_{1j}^+ + \sum_{j \in D_2} S_{2j}^-$$

Kısıtlar:

$$\sum_{i=1}^k (\lambda_i^+ - \lambda_i^-) Z_{ij} \geq d + 1 \quad j \in C_1$$

$$\sum_{i=1}^k (\lambda_i^+ - \lambda_i^-) Z_{ij} \leq d \quad j \in C_2 \quad (3.2)$$

$$\sum_{i=1}^k (\lambda_i^+ - \lambda_i^-) Z_{ij} + S_{1j}^+ - S_{1j}^- = c \quad j \in D_1$$

$$\sum_{i=1}^k (\lambda_i^+ - \lambda_i^-) Z_{ij} + S_{2j}^+ - S_{2j}^- = c \quad j \in D_2$$

$$\sum_{i=1}^k (\lambda_i^+ + \lambda_i^-) = 1$$

$$d \leq c \leq d + 1$$

$S_{1j}^+, S_{1j}^-, S_{2j}^+, S_{2j}^-, \lambda_i^+, \lambda_i^- \geq 0$ ,  $c$  ve  $d$  işaretçe serbest

(3.1) modelinde ilk aşamada doğru sınıflandırılmış tüm birimler  $\sum_{i=1}^k (\lambda_i^+ - \lambda_i^-) Z_{ij} \geq d + 1$ ,  $j \in C_1$  ve

$$\sum_{i=1}^k (\lambda_i^+ - \lambda_i^-) Z_{ij} \leq d, \quad j \in C_2 \text{ ile sınırlanmıştır.}$$

Bu kısıtlarla ilgili tüm sapma değişkenleri modelden çıkartılmıştır.  $d + 1$  ( $C_1$  alt grubu için ayırma skoru) ve  $d$  ( $C_2$  alt grubu için ayırma skoru) kısıtları altında, yeni  $c$  ayırma skoru, çakışma durumundaki gözlemlerin sapmalarının minimum yapılması ile belirlenir.  $c^*$  ve  $\lambda_i^*$  ( $= \lambda_i^+ - \lambda_i^-$ ) sırasıyla modelden elde edilen ayırma skoru ve ağırlık katsayıları olsunlar. İlk aşamada çakışma grubunda tanımlanan  $m$ . birim, aşağıdaki kurala göre sınıflandırılabilir:

$\sum_{i=1}^k \lambda_i^* Z_{im} > c^*$  ise birimin  $G_1$  grubuna ait olduğu sonucuna varılır.

$\sum_{i=1}^k \lambda_i^* Z_{im} < c^*$  ise birimin  $G_2$  grubuna ait olduğu sonucuna varılır.

Pendharkar (2011), iki gruplu sınıflandırma problemini grup 1'deki birimleri kabul sınıfı grubu, grup 2'deki birimleri de red sınıfı grubu (ya da tam tersi) kabul ederek, kabul ve red sınıfları için radyal tabanlı sinir ağına ve VZA yöntemine dayalı yeni bir yaklaşım önermiştir. Önerilen bu model İngilizce baş harflerinin kısaltması olarak RBFN-DEA olarak isimlendirilmiş ve iki gruplu sınıflandırma problemlerinin çözümünde kullanılmaktadır. Pendharkar (2011) tarafından önerilen, kabul (grup 1) ve red (grup 2) sınıfları için VZA tabanlı modeller sırasıyla (3.3) ve (3.3) modellerinde verilmektedir ve bu modeller her birim için ayrı ayrı çözülmektedir:

$$\text{Max } \zeta^t \quad ; \quad t = 1, \dots, k$$

Kısıtlar:

$$\sum_{j=1}^n \lambda_j Z_{ij} - \zeta^t Z_i^t \geq 0 \quad ; \quad i = 1, \dots, k; \quad j = 1, \dots, n$$

(3.3)

$$\sum_{j=1}^n \lambda_j = 1$$

$$\zeta^t, \lambda_j \geq 0 \quad ; \quad i = 1, \dots, k; \quad j = 1, \dots, n$$

$$\text{Min } \zeta^t \quad ; \quad t = 1, \dots, k$$

Kısıtlar:

$$\sum_{j=1}^n \lambda_j Z_{ij} - \zeta^t Z_i^t \leq 0 \quad ; \quad i = 1, \dots, k; \quad j = 1, \dots, n \quad (3.4)$$

$$\sum_{j=1}^n \lambda_j = 1$$

$$\zeta^t, \lambda_j \geq 0 \quad ; \quad i = 1, \dots, k; \quad j = 1, \dots, n$$

Bu modellerde  $t$  indisi ile ilgilenilen birim gösterilmektedir.

(3.3) modeli Grup 1'deki (kabul sınıfı) her bir KVB için çözülür ve referans kümesi belirlenir. Grup 1 (kabul sınıfı) referans kümesi  $p \leq k$  olmak üzere  $\zeta^* = 1$  olan  $p$  tane KVB'nin oluşturduğu alt kümeden meydana gelmektedir. Grup 1'deki bir birim için (3.3) modeli çalıştırılır ve bu birim referans kümesini değiştirirse, birim red sınıfına sınıflandırılır değilse kabul sınıfına sınıflandırılır. Bu adımlar grup 1'deki tüm birimler için tekrarlanır.

(3.4) modeli Grup 2'deki (red sınıfı) her bir KVB için çözülür ve referans kümesi belirlenir. Grup 2  $p \leq k$  olmak üzere  $\zeta^* = 1$  olan  $p$  tane KVB'nin oluşturduğu alt kümeden meydana gelmektedir. Bu KVB'ler referans kümesine eklenir. Grup 2'deki bir birim için (3.4) modeli çalıştırılır ve referans kümesini değiştirirse, birim grup 1'e sınıflandırılır değilse grup 2'ye sınıflandırılır. Bu adımlar grup 2'deki tüm birimler için tekrarlanır.

(3.3) ve (3.4) modeli uygulandıktan sonra bazı birimlerin çakışma durumunda olduğu yani grup 1'e mi yoksa grup 2'ye mi sınıflandırılacağı belirlenememektedir. Çakışma bölgesindeki bir birimin sınıflandırılması için, Pendharkar (2011) en yakın komşuluk özelliğinden yararlanılmasını önermiştir.  $z$  çakışma bölgesindeki bir birimin değişken vektörünü,  $z^R$  ve  $z^A$  ise sırasıyla red ve kabul sınıfı gruplarını göstermek üzere, eğer  $\|z - z^A\| > \|z - z^R\|$  ise grup 1'e,  $\|z - z^R\| > \|z - z^A\|$  ise grup 2'ye sınıflandırma yapılır. Burada  $\|\cdot\|$  ilgili vektörün Öklidyen normunu simgelemektedir.

Pendharkar ve Troutt (2014) çakışma bölgesinde yer alan ve sınıflandırılması zor olan birimleri belirlemek için Pendharkar (2011) tarafından önerilen ve (3.3)-

(3.4) ile verilen modellerine dayanan yeni bir yaklaşım geliştirmişlerdir. Bu yaklaşımda çakışma durumunda olan birimler belirlendikten sonra birimlerin skor fonksiyonlarının bilindiği varsayımı altında sınıflandırılması zor olan bu birimlerin sınıflandırılması interaktif bir yaklaşımla gerçekleştirilmektedir. Pendharkar ve Troutt (2014) çalışmasında bir lisansüstü programa başvurunun kabul edilip edilemeyeceği ile ilgili bir uygulama üzerinden interaktif yaklaşımı sunmuşlardır ve  $z_1$  lisansüstü yönetim kabul testi skoru (GMAT skoru),  $z_2$  lisans ortalaması (GPA skoru) olmak üzere  $u(z) = z_1 + 200z_2$  gibi bir skor fonksiyonunun bilindiğini varsaymaktadırlar. Okul yöneticileri skor değerinin 1000'den çok ve az oluşuna göre lisansüstü programa kabul edilip edilmemesine karar vermektedirler. Yani yöneticiler bir bakıma kendi koydukları bir sınıflandırma fonksiyonu ile programa kabul ve red gruplarını oluşturmaktadırlar. Pendharkar ve Troutt (2014) çalışmasında grupların oluşturulmasını sağlayan bu türden bir kuralın bilindiği varsayılmaktadır. Pendharkar ve Troutt (2014) çalışmasının temelinde (3.3) ve (3.4) modelleri ile belirlenen grup 2 (red sınıfı) referans sınırının grup 1 (kabul sınıfı) örneklerinden, grup 1 (kabul sınıfı) referans sınırının ise grup 2 (red sınıfı) örneklerinden etkilenip etkilenmemesine göre çakışma durumunda olan örneklerin belirlenmesi prensibi yatmaktadır.  $A$  kümesi grup 1 (kabul sınıfı) örneklerinin kümesini,  $R$  kümesi grup 2 (red sınıfı) örneklerinin kümesini,  $A^R$  kümesi grup 1 (kabul sınıfı) referans kümesini ve  $R^R$  kümesi grup 2 (red sınıfı) referans kümesini gösterebilir. Burada  $A^R$  ve  $R^R$  ile gösterilen referans kümelerinin (3.3) ve (3.4) ile belirlendiği aşikâr bir durumdur. İlk olarak  $A^R$  grup 1 (kabul sınıfı) referans kümesi ile grup 2 (red sınıfı) örnekleri birleştirilir ve (3.3) ile verilen model  $A^R \cup R$  kümesine uygulanır. Eğer bir grup 2 (red sınıfı) örneği orijinal  $A^R$  grup 1 (kabul sınıfı) referans kümesini değiştirirse yani bir grup 2 (red sınıfı) örneği  $\zeta^* = 1$  optimum amaç fonksiyon değerine sahip olursa bu örnek grup 2 (kabul sınıfı) referans kümesinin altında bir konumdadır ve bu durumda olan örnekler grup 2'ye (red sınıfına) sınıflandırılır. Benzer şekilde  $R^R$  grup 2 (red sınıfı) referans kümesi ile grup 1 (kabul sınıfı) örnekleri birleştirilir ve (3.4) ile verilen model  $R^R \cup A$  kümesine uygulanır. Eğer bir grup 1 (kabul sınıfı) örneği orijinal  $R^R$  grup 2 (red sınıfı) referans kümesini değiştirirse yani bir kabul sınıfı örneği  $\zeta^* = 1$  optimum amaç fonksiyon değerine sahip olursa bu örnek grup 2 (red sınıfı) referans kümesinin üstünde bir konumdadır ve bu durumda olan örnekler grup 1'e (kabul sınıfına) sınıflandırılır. Sınıflandırılması yapılamayan örnekler yani çakışma durumunda olan örnekler grup 1 (kabul sınıfı) referans sınırı (kümesi)

ve grup 2 (red sınıfı) referans sınırı (kümesi) birimlerinin oluşturduğu konveks bir küme ile çevrelenmektedir.

#### 4. Önerilen Model

Sueyoshi (1999, 2004) DEA-DA ve genişletilmiş DEA-DA yaklaşımlarında, ilk aşamada çakışma durumunun tespit edildiği ikinci aşamada ise çakışma durumunda olan birimlerin sınıflandırıldığı iki aşamalı iki farklı model önermiştir. Sueyoshi (1999, 2004) DEA-DA çalışmalarından esinlenerek Pendharkar ve Troutt (2014) modeli ile sınıflandırılmayan birimlerin detaylı bir inceleme ile sınıflandırılması (3.5) ile verilen modelde gösterilmektedir.

(3.3) ve (3.4) modeli ile sırasıyla  $A^R$  ve  $R^R$  kabul ve red sınıfı referans kümeleri belirlenir.  $\lambda_i^* = 1$  ve  $\zeta^* = 1$  (3.3) ve (3.4) modellerinden elde edilen optimum çözümler olsunlar. Değeri  $z_m$  ile gösterilmiş  $m$ . birimin  $G_1$ ,  $G_2$  ya da  $G_1 \cap G_2$  (çakışma) durumlarından hangisine ait olduğu aşağıda verilen kriterle belirlenir.

İlk olarak  $A^R$  grup 1 (kabul sınıfı) referans kümesi ile grup 2 (red sınıfı) örnekleri birleştirilir ve (3.3) ile verilen model  $A^R \cup R$  kümesine uygulanır. Eğer grup 2'deki bir birim orijinal  $A^R$  kabul sınıfı referans kümesini değiştirirse yani bir grup 2 birimi (red sınıfı örneği)  $\zeta^* = 1$  optimum amaç fonksiyon değerine sahip olursa bu örnek grup 1 (kabul sınıfı) referans kümesinin altında bir konumdadır ve bu durumda olan örnekler grup 2'ye (red sınıfına) sınıflandırılır. Benzer şekilde  $R^R$  grup 2 (red sınıfı) referans kümesi ile grup 1 (kabul sınıfı) örnekleri birleştirilir ve (3.4) ile verilen model  $R^R \cup A$  kümesine uygulanır. Eğer grup 1'deki bir birim (bir kabul sınıfı örneği) orijinal  $R^R$  grup 2 (red sınıfı) referans kümesini değiştirirse yani grup 1'deki bir birim (bir kabul sınıfı örneği)  $\zeta^* = 1$  optimum amaç fonksiyon değerine sahip olursa bu örnek grup 2 (red sınıfı) referans kümesinin üstünde bir konumdadır ve bu durumda olan örnekler grup 1'e (kabul sınıfına) sınıflandırılır.

$A^R \cup R$  kümesine (3.3) modeli uygulanır. Eğer grup 2'deki bir birim (bir red sınıfı örneği) orijinal  $A^R$  grup 1 (kabul sınıfı) referans kümesini değiştirirse bu örnek grup 2'ye (red sınıfına) sınıflandırılır. Bu sınıfı  $C_1$  ile gösterelim.

$R^R \cup A$  kümesine (3.4) modeli uygulanır. Eğer grup 1'deki bir birim (bir kabul sınıfı örneği) orijinal  $R^R$  grup 2 (red sınıfı) referans kümesini değiştirirse bu

örnek grup 1'e (kabul sınıfına) sınıflandırılır. Bu sınıfı  $C_2$  ile gösterelim.

$D_1 = G_1 - C_1$  ve  $D_2 = G_2 - C_2$  sınıfları ise çakışma durumundaki örneklerden oluşmaktadır.

Çakışmanın varlığı tanımlandıktan sonra, bir yeni ayırma skoru ( $c$ ) modelin içine sadece  $D_1 \cup D_2$  alt grubunda yer alan birimleri sınıflandıracak şekilde konulur. Bu modelde  $C_1$  ve  $C_2$ 'deki birimlerin sırasıyla grup 1 ve grup 2'de yani kendi gruplarında kalmaları bu birimler için oluşturulan kısıtlara herhangi bir sapma değişkeni koymayarak sağlanmaktadır.

$$\text{Min} \sum_{j \in D_1} S_{1j}^+ + \sum_{j \in D_2} S_{2j}^-$$

Kısıtlar:

$$\sum_{i=1}^k \lambda_i Z_{ij} \geq \delta \quad j \in C_1$$

$$\sum_{i=1}^k \lambda_i Z_{ij} \leq 0 \quad j \in C_2 \quad (3.5)$$

$$\sum_{i=1}^k \lambda_i Z_{ij} + S_{1j}^+ - S_{1j}^- = c \quad j \in D_1$$

$$\sum_{i=1}^k \lambda_i Z_{ij} + S_{2j}^+ - S_{2j}^- = c \quad j \in D_2$$

$$\sum_{i=1}^k \lambda_i = 1$$

$\delta$  küçük pozitif bir sabit,  $S_{1j}^+, S_{1j}^-, S_{2j}^+, S_{2j}^-, \lambda_i \geq 0, c$  işaretçe serbest

## 5. Açıklayıcı Örnek

Araştırma önerisi kısmında, Pendharkar ve Troutt (2014)'den alınan bir örnek üzerinde önerilen yöntemin işleyişi ve ayrıca Pendharkar ve Troutt (2014) modeli ve Sueyoshi (2004) iki aşamalı VZA sınıflandırma modeli ile görsel bir karşılaştırmaya da yer verilmiştir.

Bir lisansüstü programa başvurunun kabul edilip edilemeyeceği ile ilgili olan bu örnekte  $z_1$  lisansüstü yönetim kabul testi skoru (GMAT skoru) ve  $z_2$  lisans ortalaması (GPA skoru) değerini göstermektedir. Veri kümesi Tablo 1'de verilmektedir. Kabul sınıfında 16, red sınıfında ise 13 birim vardır. Birimlerin sıra

numarasına göre  $A = \{1, 2, 3, \dots, 14, 15, 16\}$  ve  $R = \{17, 18, 19, \dots, 27, 28, 29\}$  kümeleri oluşur.

Yöntemin uygulamasında ilk olarak (3.3) ve (3.4) modelleri grup 1 (kabul sınıfı) referans kümesi ( $A^R$ ) ve grup 2 (red sınıfı) referans kümesi ( $R^R$ ) belirlenir. Modellerin çözümünden grup 1'i zarflayan kabul sınıfı referans kümesi  $A^R = \{4, 11, 15, 16\}$  ve grup 2'yi zarflayan red sınıfı referans kümesi  $R^R = \{25, 26, 29\}$  şeklinde elde edilir.

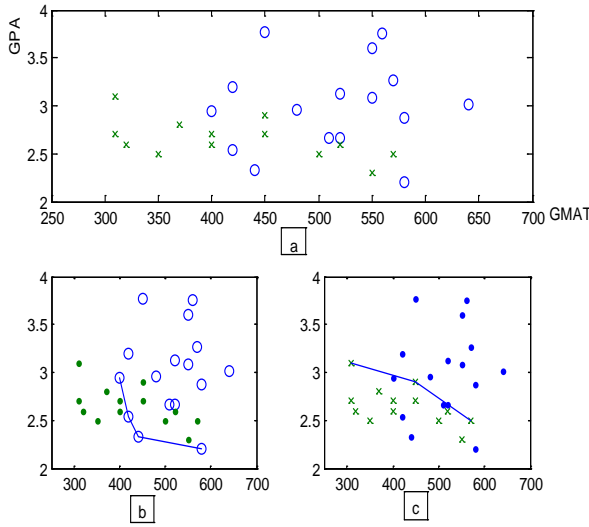
**Tablo 1.** Bir lisansüstü başvuruya ilişkin veriler

Başvuru numarası	GMAT	GPA	Karar
1	640	3,02	Kabul
2	550	3,09	Kabul
3	510	2,67	Kabul
4	420	2,54	Kabul
5	560	3,75	Kabul
6	550	3,60	Kabul
7	580	2,87	Kabul
8	420	3,20	Kabul
9	450	3,77	Kabul
10	520	2,67	Kabul
11	440	2,33	Kabul
12	480	2,96	Kabul
13	520	3,13	Kabul
14	570	3,26	Kabul
15	400	2,95	Kabul
16	580	2,20	Kabul
17	310	2,70	Red
18	350	2,50	Red
19	400	2,70	Red
20	370	2,80	Red
21	450	2,70	Red
22	500	2,50	Red
23	520	2,60	Red
24	550	2,30	Red
25	570	2,50	Red
26	450	2,90	Red
27	320	2,60	Red
28	400	2,60	Red
29	310	3,10	Red

İlk olarak  $A^R \cup R = \{4, 11, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29\}$  veri kümesinden oluşan birimlere (3.3) ile verilen model uygulanır. 17, 18, 19, 20, 25, 26, 27, 28 ve 29



numaralı red sınıfı birimleri  $\zeta^* = 1$  amaç fonksiyon değerine sahip olduklarından ve bu yüzden de orijinal  $A^R$  referans kümesini değiştiren birimler olduklarından kabul sınıfı referans kümesinin altında olduklarına karar verilir ve bu birimler grup 2'ye (red sınıfına) sınıflandırılırlar. Benzer şekilde  $R^R \cup A = \{25, 26, 29, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16\}$  veri kümesinden oluşan birimlere (3.4) ile verilen model uygulanır. 3 numaralı birim dışındaki birimler  $\zeta^* = 1$  amaç fonksiyon değerine sahip olduğundan ve bu yüzden de orijinal  $R^R$  referans kümesini değiştiren birimler olduğundan red sınıfı referans kümesinin altında olduğuna karar verilir ve bu birimler grup 1'e (kabul sınıfına) sınıflandırılırlar.



**Şekil 1.** a) Verilerin iki boyutlu uzayda gösterimi, b) Kabul sınıfı için sınır, c) Red sınıfı için sınır

Geriyeye kalan 3, 4, 11, 15, 16, 21, 22, 23, 24, 25, 26 ve 29 numaralı birimler kesin olarak hangi sınıfa ait olduğu belirlenemeyen birimlerdir ve çakışma durumunda olan birimlerdir.

Şekil 1a, verilerin iki boyutlu uzayda gösterimini, Şekil 1b, Kabul sınıfı için referans kümesini Şekil 1c ise red sınıfı için referans kümesini göstermektedir.

(3.3) ve (3.4) modelinin sonuçlarına göre; 17, 18, 19, 20, 25, 26, 27, 28 ve 29 numaralı red sınıfı birimleri grup 2'ye (red sınıfına) ve 1, 2, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15 ve 16 numaralı kabul sınıfı birimleri grup 1'e sınıflandırılmaktadır. Geriyeye kalan yani çakışma durumunda olan 3, 4, 11, 15, 16, 21, 22, 23, 24, 25, 26 ve 29 numaralı birimler için (3.5) ile verilen model çözülerek nihai sınıflandırma yapılmaktadır. (3.3) ve (3.4) modellerli sonuçlarına göre doğru gruplarına atanan birimler (3.5) modelinde herhangi bir sapma değişkeni olmadan yer almaktadırlar ve bu sayede doğru grupları (3.5) modeli sonucunda değişmemektedir. (3.5) modeli ile çakışma

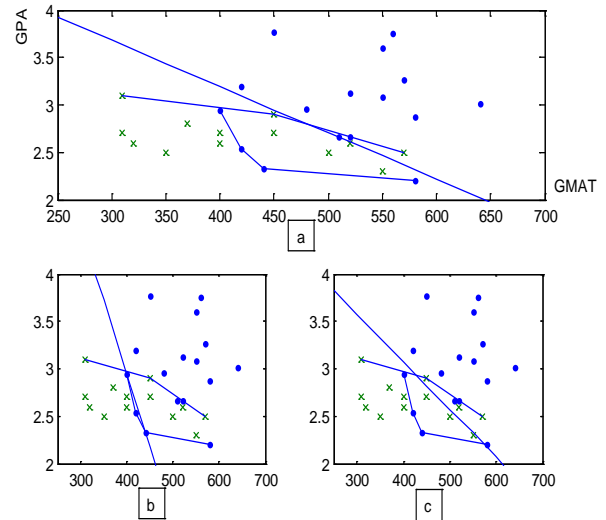
durumunda olan birimler mümkün olduğunca doğru sınıflandırılmaya çalışılmaktadır.

Pendharkar skor değerlerinin bilinmediği durumlarda literatürde önerilen bir matematiksel programlama yaklaşımının kullanılabilirliğini önermiştir ve Silva ve Stam (1997) modelini tüm birimlere uygulayarak

$0.645161z_1 + 0.01z_2 - 5.90323$  sınıflandırma fonksiyonunu elde etmiştir ve bu fonksiyon yardımıyla 22 birim doğru sınıflandırılabilmiştir. Bu yaklaşımda grup 1 referans kümesinin altında kalan ya da grup 2 referans kümesinin üstünde kalan birimlerin kendi gruplarında kalması korunmadan, doğrudan tüm birimlere iki grubu ayırmak için bir matematiksel programlama yaklaşımı kullanılmaktadır. Sueyoshi'nin iki aşamalı modelinin sonucunda

$0.92593z_1 + 0.004633z_2 - 4.69908$  sınıflandırma fonksiyonuna ulaşılmıştır ve bu fonksiyon yardımıyla 23 birim doğru sınıflandırılabilmiştir. Pendharkar (2014) ve Sueyoshi (2001) yaklaşımlarının bir karması olarak düşünülebilecek olan yeni yaklaşımla ise  $1.07148z_1 + 0.005215z_2 - 5.50847$  sınıflandırma fonksiyonu elde edilmiştir ve bu fonksiyon yardımıyla 24 birim doğru sınıflandırılabilmiştir.

Şekil 2a'da, önerilen yöntemle elde edilen ve 24 birimin doğru sınıflandırıldığı ayırma sınırının gösterimi, Şekil 2b'de, Pendharkar (2014) yöntemiyle elde edilen ve 21 birimin doğru sınıflandırıldığı ayırma sınırının gösterimi, Şekil 2c'de ise Sueyoshi (2004) iki aşamalı yöntemiyle elde edilen ve 23 birimin doğru sınıflandırıldığı ayırma sınırının gösterimi yer almaktadır.



**Şekil 2.** a) Önerilen yöntemle elde edilen ayırma sınırının gösterimi, b) Pendharkar (2014) yöntemiyle elde edilen ayırma sınırı, c) Sueyoshi (2004) iki aşamalı modeli elde edilen ayırma sınırı

## 6. Simülasyon Çalışması

Bu bölümde, iki gruplu durum için Fisher'in doğrusal ayırma fonksiyonu, Pendharkar (2014) tarafından önerilen model, Sueyoshi (2001) tarafından önerilen model ve Pendharkar (2014) ve Sueyoshi (2001) modellerinden esinlenerek önerilen yeni model, geri yayılım algoritması ile eğitilmiş çok katmanlı ağ yapısı yaklaşımlarının sınıflandırma performanslarını incelemek için, bir 1000 tekrarlı Monte Carlo simülasyon çalışması tasarlanmıştır.

Sınıflandırma yöntemlerinin performanslarını karşılaştırmak için çapraz geçerlilik (cross validation) teknikleri kullanılmaktadır. Literatürde en çok kullanılan çapraz geçerlilik teknikleri; doğrulama örneği çapraz geçerlilik (test-holdout sample cross validation), bir birimi dışarıda tutma çapraz geçerlilik (leave-one-out cross validation), bazı birimleri dışarıda tutma çapraz geçerlilik (leave-some-out cross validation) ve  $k$  – kat çapraz geçerlilik ( $k$  – fold cross validation) yaklaşımlarıdır.

Doğrulama örneği çapraz geçerlilik yaklaşımında veri seti rasgele olarak iki sete ayrılır. Setlerden biri ayırma fonksiyonunun bulunuşunda kullanılır ve bu set eğitim veya geliştirme örneği (training sample, development sample) olarak isimlendirilmektedir. Diğer set ise doğrulama örneği (test-holdout sample) olarak isimlendirilir ve bu örnekler ayırma fonksiyonunun bulunuşunda kullanılmaz. Eğitim örneğinde ayırma fonksiyonu elde edildikten sonra yöntemin doğru sınıflandırma performansı doğrulama örneği ile sınanır.

Bu çalışmada, doğrulama örneği çapraz geçerlilik yaklaşımı kullanılmıştır.

Simülasyon çalışmasında, iki grup için altı farklı veri tipinde üç bağımsız değişkenli Düzgün (Uniform) dağılımdan veri üretilmiştir. Bu veri tiplerinden üç tanesi (B, D ve F veri tipleri) aykırı değerler tarafından bozulmuş veri tipleridir. Çalışmada kullanılan veri tipleri Tablo 2'de özetlenmiştir.

**Tablo 2.** Simülasyon çalışmasında kullanılan veri tipleri

Veri Tipi	Grup 1	Grup 2
A	$\{U[5,9], U[5,9], U[5,9]\}$	$\{U[3,7], U[3,7], U[3,7]\}$
B	%90 $\{U[5,9], U[5,9], U[5,9]\}$ %10 $\{U[3,5], U[3,5], U[3,5]\}$	%90 $\{U[3,7], U[3,7], U[3,7]\}$ %10 $\{U[7,9], U[7,9], U[7,9]\}$
C	$\{U[5,9], U[5,9], U[5,9]\}$	$\{U[4,8], U[4,8], U[4,8]\}$

D	%90 $\{U[5,9], U[5,9], U[5,9]\}$ %10 $\{U[3,5], U[3,5], U[3,5]\}$	%90 $\{U[4,8], U[4,8], U[4,8]\}$ %10 $\{U[8,10], U[8,10], U[8,10]\}$
E	$\{U[3,11], U[3,11], U[3,11]\}$	$\{U[4,8], U[4,8], U[4,8]\}$
F	%90 $\{U[3,11], U[3,11], U[3,11]\}$ %10 $\{U[1,3], U[1,3], U[1,3]\}$	%90 $\{U[4,8], U[4,8], U[4,8]\}$ %10 $\{U[8,10], U[8,10], U[8,10]\}$

Çok katmanlı ağ geri yayılım algoritması ile eğitilirken beş gizli katmanı olan ağ yapısı kullanılmıştır. Maksimum iterasyon sayısı 10000 olarak öğrenme oranı ise 0,1 olarak seçilmiştir.

Toplam olarak  $n = 200$  birim ve gruplarda da 100'er birim olduğu varsayılmıştır. İki grup için 100'lük birimlere ilişkin özellikler Tablo 2'deki veri tiplerinden rasgele olarak üretilmiştir. Toplam 200 birimin 100 tanesi eğitim örneği (training sample), geriye kalan 100 tanesi de doğrulama örneği (holdout sample) olarak kullanılmıştır. Tablo 3'de  $n_1 = n_2 = 50$  durumunda, ilgili veri tipinden birinci gruba 100 ve ikinci gruba 100 olmak üzere 200 birim alınmış, birinci ve ikinci gruptaki 100'er birimin 50 tanesi eğitim örneği, 50 tanesi de tahmin örneği olarak kullanılmıştır.

Tablo 3 incelendiğinde önerilen iki gruplu sınıflandırma modelinin, bütün veri tiplerinde çok büyük çoğunlukla diğer 3 yaklaşımdan daha büyük doğru sınıflandırma oranına sahip olduğu görülmektedir.

**Tablo 3.** Tahmin örneklerdeki 5 yaklaşıma ilişkin doğru sınıflandırma oranları ( $n_1 = n_2 = 50$ )

Veri Tipi	FLDF	Pendharkar (2014)	Önerilen model	Sueyoshi	YSA GY
A	83.81 (4.69)*	84.88 (4.16)	86.42 (4.76)	85.12 (3.99)	85.16 (4.52)
B	81.13 (4.71)	81.93 (5.11)	85.51 (4.59)	85.05 (4.39)	85.05 (4.14)
C	83.29 (4.51)	83.05 (4.97)	85.87 (5.41)	85.48 (4.67)	85.12 (5.14)

D	81.71 (5.19)	82.92 (4.95)	83.11 (5.17)	83.10 (5.41)	83.07 (4.95)
E	82.31 (4.29)	84.89 (5.32)	87.32 (4.41)	86.68 (4.39)	87.26 (5.35)
F	80.76 (5.07)	83.56 (4.89)	85.40 (4.98)	84.21 (4.26)	83.96 (4.88)

\* 1000 tekrardaki doğru sınıflandırma sayılarına ilişkin standart sapmalar

Bu kesimde Tablo 3'den elde edilen sonuçlar kullanılarak önerilen modelin FLDF ve diğer yöntemlerle doğru sınıflandırma performansları eşleştirme  $t$  testi yardımıyla karşılaştırılmaktadır.

Hipotez testleri 6 veri tipi durumu (A, B, C, D, E ve F), geliştirme örneği ve tahmin örneği olarak seçilen  $n_1 = n_2 = 50$  için tekrarlanmıştır. Hipotezler şu şekilde oluşturulmuştur:

$H_0$ : Önerilen modelin doğru sınıflandırma ortalaması ile  $i$ . yöntemin doğru sınıflandırma ortalaması arasında fark yoktur.

$H_1$ : Önerilen modelin doğru sınıflandırma ortalaması  $i$ . yöntemin doğru sınıflandırma ortalamasından büyüktür ( $i=FLDF, Pendharkar (2014), Sueyoshi, YSA-geri yayılım$ )

Örneğin, Tablo 3'de, A veri tipi (grup 1;  $\{U[5,9], U[5,9], U[5,9]\}$ , grup 2;  $\{U[3,7], U[3,7], U[3,7]\}$ ) durumundaki FLDF modeline ait olan 83,81 ve 4,69 değerleri, sırasıyla 1000 tekrardaki (örnekteki) FLDF modelinin doğru sınıflandırma ortalaması ve standart sapmasıdır. Önerilen modele ait 86,42 ve 4,76 değerleri de sırasıyla 1000 örnekteki doğru sınıflandırma ortalaması ve standart sapmasıdır. Bu bilgiler kullanılarak aşağıdaki hipotez testi yapılmaktadır.

$H_0$ : Önerilen modelin doğru sınıflandırma ortalaması ile FLDF'nin doğru sınıflandırma ortalaması arasında fark yoktur.

$H_1$ : Önerilen modelin doğru sınıflandırma ortalaması FLDF'nin doğru sınıflandırma ortalamasından büyüktür.

Tablo 4'de geri yayılım algoritması ile eğitilen çok katmanlı ağ yapısının doğru sınıflandırma ortalamasının, FLDF, Pendharkar (2014), Sueyoshi (2001) ve YSA-Geri Yayılım yöntemlerinin doğru sınıflandırma ortalamasından büyük olduğunu iddia eden hipotez testinin sonuçları yer almaktadır.

Tablo 4'deki  $n_1 = n_2 = 50$  durumundaki 15,23<sup>a</sup> değeri, A veri tipi durumunda, geri yayılım algoritması ile eğitilen çok katmanlı ağ yapısının doğru sınıflandırma ortalamasının FLDF yönteminin doğru sınıflandırma ortalamasından büyük olduğunun testindeki,  $t$  test istatistiğinin aldığı değeri gösterir. <sup>a</sup> ise bu hesaplanan 14,12'ün  $p$  değeridir ve  $p$  değeri de 0,01'den küçük bir değerdir. Test istatistiğinin aldığı 14,12 değerine göre veya  $p$  değerine göre hipotezin sonucuna karar verilebilir.  $p$  değerinin 0,01'den küçük olması 0,01 anlamlılık düzeyinde  $H_0$ 'ın reddedildiği sonucunu çıkarır. Sonuçta, istatistiksel olarak geri yayılım algoritması ile eğitilen çok katmanlı ağ yapısının doğru sınıflandırma ortalaması FLDF'nin doğru sınıflandırma ortalamasından büyük olduğu hipotezi kabul edilmiştir.

Tablo 4'de elde edilen sonuçlar incelendiğinde önerilen modelin FLDF ve Pendharkar (2014) yaklaşımlarından D veri seti haricinde her durumda üstün olduğu gözlenmektedir. Sueyoshi (2001) ve geri yayılım algoritması ile eğitilen çok katmanlı ağ yapısı ile karşılaştırıldığında ise tüm veri setlerinde performans değeri olarak üstün, A ve F veri setlerinde de istatistiksel olarak üstündür.

İki gruplu durum için, simülasyon çalışması sonuçlarına göre önerilen modelin sınıflandırma problemlerinde başarı ile kullanılacağı sonucuna varılabilir.

**Tablo 4.** Doğrulama örneklerindeki önerilen model diğer (FLDF, Pendharkar (2014), Sueyoshi (2001), geri yayılım algoritması ile eğitilmiş çok katmanlı ağ yapısı), karşı hipotez testi sonuçları (eşleştirme  $t$  değerleri)

Test 1 (FLDF)		Test 3 (Sueyoshi)	
A	14.12 <sup>a</sup>	A	3.01 <sup>a</sup>
B	15.11 <sup>a</sup>	B	1.01
C	8.34 <sup>a</sup>	C	1.41
D	7.53 <sup>a</sup>	D	0.19
E	20.23 <sup>a</sup>	E	2.01
F	17.29 <sup>a</sup>	F	3.25 <sup>a</sup>
Test 2 (Pendharkar)		Test 4 (YSA-GY)	
A	3.32 <sup>a</sup>	A	3.79 <sup>a</sup>
B	5.01 <sup>a</sup>	B	1.32

<b>C</b>	3.22 <sup>a</sup>	<b>C</b>	1.99
<b>D</b>	1.32	<b>D</b>	0.11
<b>E</b>	3.12 <sup>a</sup>	<b>E</b>	0.99
<b>F</b>	2.88 <sup>a</sup>	<b>F</b>	2.98 <sup>a</sup>

<sup>a</sup>  $H_0$  Red ( $p$  değ.  $< 0,01$ )

## 7. Sonuçlar

Bu çalışmada iki grup sınıflandırma problemini çözmede VZA tabanlı Sueyoshi (2004) ve Pendharkar (2014) modellerinin bir karması olan yeni bir sınıflandırma yaklaşımı geliştirilmiştir. Önerilen yaklaşım VZA-BCC modeli sayesinde parçalı doğrusal sınıflandırma düzlemleri üretebilmekte ve iki aşamalı inceleme fikri ile detaylı bir inceleme yapabilmektedir. Önerilen yaklaşımın aşamaları Pendharkar (2014)'den alınan bir problem üzerinde ayrıntılı olarak incelenmiş ve ulaşılan sonuçlar, bu yaklaşımın uygulanabilirliğini ve etkinliğini göstermiştir. Ayrıca yapılan simülasyon çalışması ile önerilen yöntemin sınıflandırma performansının literatürde önerilen yöntemlerden de daha iyi olduğu istatistiksel olarak doğrulanmıştır. İleriki çalışmalarda, önerilen yöntemin negatif değer içeren veri setlerinde ve ayrıca çok gruplu sınıflandırma problemlerindeki performansları da incelenebilir.

## Kaynaklar

Anderson, T.W., 1984. "An introduction to multivariate analysis", Wiley, New York, USA, 10-25.

Bal, H., Örkücü, H.H., 2005. "Combining the Discriminant Analysis and Data Envelopment Analysis in view of Multiple Criteria Decision Making: A New Model", *G.U. Journal of Science*, 18 (3), 355-364.

Bal, H., Örkücü, H.H., Çelebioğlu S., 2006a. "An Experimental Comparison of the New Goal Programming and Linear Programming Approaches in the Two-Group Discriminant Problems", *Computers&Industrial Engineering*, 50 (3), 296-311.

Bal, H., Örkücü, H.H., Çelebioğlu S., 2006b. "An Alternative Model to Fisher and Linear Programming Approaches in Two-Group Classification Problem: Minimizing Deviations from the Group Median", *G.U. Journal of Science*, 19 (1), 49-55.

Bal, H., Örkücü, H.H., 2007. "Data Envelopment Analysis Approach to Two-Group Classification Problems and an Experimental Comparison with Some Classification Models", *Hacettepe Journal of Mathematics and Statistics*, 36(2), 169-180.

Bal, H., Örkücü, H.H., 2011. "A new mathematical programming approach to Multi-Group Classification

Problems", *Computers and Operations Research*, 38(1), 105-111.

Banker, R.D., Charnes, A., Cooper, W.W., 1984. "Some models for estimating technical and scale inefficiencies in data envelopment analysis", *Management Science*, 30(9): 1078-1092.

Charnes, A., Cooper, W.W., Rhodes, E., 1978. "Measuring the efficiency of decision making units", *European Journal of Operational Research*, 2: 429-444.

Cooper, W.W., Seiford, L.M., Tone, K., 2000. "Data envelopment analysis", *Kluwer Academic Publishers*, Boston USA., 25-60.

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

Fred, N., Glover, F., 1981a. "A Linear programming approach to the discriminant problem", *Decision Sciences*, 12: 68-74.

Fred, N., Glover, F., 1981b. "Simple but powerful goal programming formulations for the statistical discriminant problem", *European Journal of Operational Research*, 7: 44-60

Fred, N., Glover, F., 1986a. "Evaluating alternative linear programming models to solve the two-group discriminant problem", *Decision Sciences*, 17: 151-162.

Fred, N., Glover, F., 1986b. "Resolving certain difficulties and improving the classification power of LP discriminant analysis formulations", *Decision Sciences*, 17: 589-595.

Glover, F. 1990., "Improving linear programming models for the discriminant problem", *Decision Sciences*, 21: 771-785.

Hosseini, J.H., Armacost, R.L., 1994. "Two-group discriminant problem with equal group mean vectors: An experimental evaluation of six linear/nonlinear programming formulations", *European Journal of Operational Research*, 77: 241-252.

Joachimsthaler, E.A., Stam, A., 1988. "Four approaches to the classification problem in discriminant analysis: An experimental study", *Decision Sciences*, 19: 322-333.

Koehler, G.J., 1989. "Characterization of unacceptable solutions in LP discriminant analysis", *Decision Sciences*, 21: 239-257.

Koehler, G.J., Erenguc, S.S., 1990. "Minimizing misclassifications in linear discriminant problem", *Decision Sciences*, 21: 63-85.

Markowski, E.P., Markowski, C.A., 1985. "Some difficulties and improvements in applying linear

- programming formulations to the discriminant problem", *Decision Sciences*, 16: 237-247.
- Lachenburch P.A., 1975. "Discriminant analysis", *Hafner Press*, New York, USA, 40-90.
- Lam, K.F., Choo, E.U., Moy, J.W., 1996. "Minimizing deviations from the group mean: A new linear programming approach for the two-group classification problem", *European Journal of Operational Research*, 88: 358-367.
- Lam, K.F., Moy, J.W., 1996. "Improved linear programming formulations for the multi-group discriminant problem", *Journal of Operational Research Society*, 47: 1526-1529.
- Lam, K.F., Moy, J.W., 1997. "An experimental comparison of some linear programming approaches to the discriminant problem", *Computers and Operations Research*, 24 (7), 593-599.
- Lam, K.F., Moy, J.W., 2002. "Combining discriminant method in solving classification problems in two-group discriminant analysis", *European Journal of Operational Research*, 138: 294-301.
- Lee, C.K., Ord, J.K., 1990. "Discriminant analysis using least absolute deviations", *Decision Sciences*, 21: 86-96
- .
- Örkcü, H.H., Bal, H., 2011. "A Combining Mathematical Programming Method for Multi-Group Data Classification", *G.U. Journal of Science*, 24 (1), 77-84.
- Pendharkar, P.C., 2011. "A hybrid radial basis function and data envelopment analysis neural network classification", *Computers and Operations Research*, 38 (1): 256-266.
- Pendharkar, P.C. and Troutt, M.D., 2014. "Interactive classification using data envelopment analysis", *Annals of Operations Research*, 214 (1): 125-141.
- Rubin, A., 1990. "A comparison of linear programming and parametric approaches to the two-group discriminant problem", *Decision Sciences*, 21: 373-386.
- Sueyoshi, T., 1999. "DEA-Discriminant analysis in the view of goal programming", *European Journal of Operational Research*, 115: 564-582.
- Sueyoshi, T., 2001. "Extended DEA-Discriminant analysis", *European Journal of Operational Research*, 131: 324-351.
- Sueyoshi, T., 2004. "Mixed integer programming approach of extended DEA-Discriminant analysis", *European Journal of Operational Research*, 152: 45-55.
- Stam, A., Jones, D.G., 1990. "Classification performance of mathematical programming techniques in discriminant analysis: Results for small and medium sample sizes", *Managerial and Decision Economics*, 11: 243-253.
- Stam, A., Ragsdale, C.T., 1992. "On the classification gap in mathematical programming based approaches to the discriminant problem", *Naval Research Logistic*, 39: 545-559.