Veri madenciliğinde yeni çok-sınıflı sınıflandırma yöntemleri ve şirketlerin finansal özelliklerine göre derecelendirilmesi problemine uygulanması

Proje No: 107M472

Doç.Dr. Refail KASIMBEYLİ
Yrd. Doç.Dr. Gürkan ÖZTÜRK
Yrd. Doç.Dr. Özden ÜSTÜN

OCAK 2010

İZMİR

Önsöz

Veri madenciliğinde sınıflandırma problemleri günümüzde birçok alanda geniş bir uygulama yelpazesine sahip olmaktadır. Sınıflandırma problemlerinin günlük hayatımızda yapılan uygulamalardaki önemi de giderek artmaktadır.

Bu sebeptendir ki, sınıflandırma problemleri ve yeni sınıflandırma yöntemlerinin geliştirilmesi üzerine yapılan araştırmalar da bütün dünyada giderek yoğunlaşmaktadır.

Hiç şüphesizdir ki, kesin çözümlerin bulunmasına dayalı matematiksel programlama yaklaşımları bu çalışmalarda daha önemli bir yer tutmaktadır.

Biz de bu projede, matematiksel programlama ve genellikle dışbükey olmayan analiz alanında son zamanlarda geliştirdiğimiz özel yaklaşımları sınıflandırma problemlerinin araştırılmasında kullanarak yeni yöntemler geliştirdik. Bu yöntemleri kullanarak bütün dünyada bilinen test problemleri üzerinde yaptığımız denemelerde oldukça iyi sonuçlar elde ettik. Bu yöntemleri İMKB'de işlem gören şirketlerin derecelendirilmesi problemine de uyguladık. Elde edilen başarılı sonuçlar, geliştirmiş olduğumuz yöntemlerin etkinliğini çok güzel ortaya koymaktadır.

Bu proje kapsamında yapılan çalışmalar, burada başlatılan araştırmaların başarılı bir şekilde ileride yapılacak birçok araştırmaya ışık tutacak nitelikte olduğunu göstermektedir.

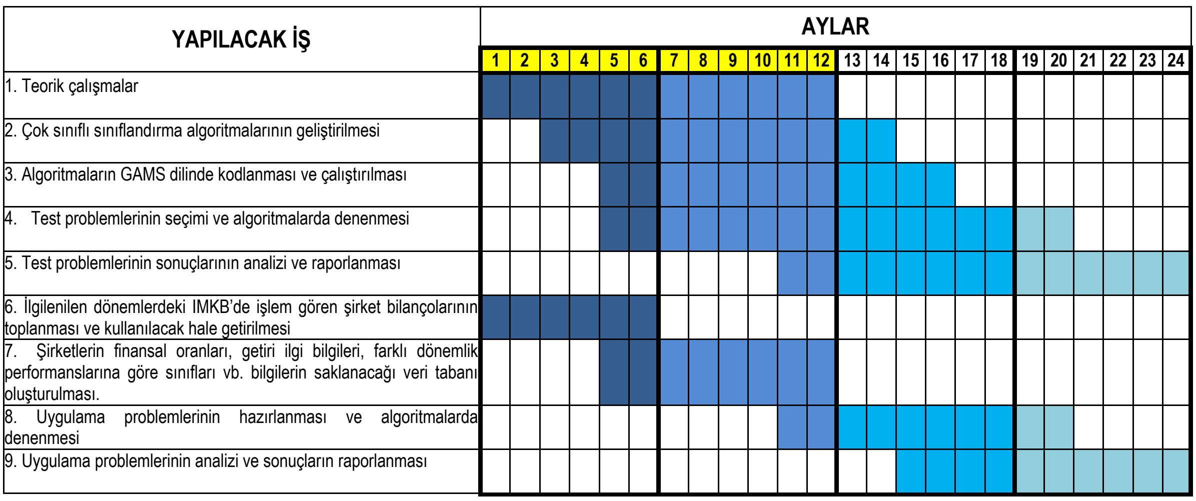
İster bu proje kapsamında geliştirilen yeni yöntemler olsun, isterse de yine de bu proje kapsamında çözüm getirmeye çalıştığımız ve ilk defa proje ekibi tarafından tasarlanmış olan İMKB'de işlem gören şirketlerin derecelendirilmesi problemi, hem problemlerin tanımlanması açısından hem de çözümleri açısından, oldukça zor problemlerdir. Bu problemlerin çözümü, TÜBİTAK desteği olmadan gerçekleştirilemezdi. Bu desteğinden dolayı TÜBİTAK'a ve araştırma projelerinin desteklenmesi politikalarının oluşturulmasında emeği geçen herkese teşekkür ediyoruz. Projenin gerçekleştirilmesi sürecinde yakın desteğini esirgemeyen ve projeye ev sahipliği yapan İzmir Ekonomi Üniversitesine de teşekkür ediyoruz.

Refail Kasımbeyli, Gürkan Öztürk, Özden Üstün.

İzmir, Ocak 2010.

İçindekiler

Şekil İndeksi



Tablo İndeksi

Özet

Bu projede, iki ve daha fazla sınıfa sahip gerçek hayat veri kümelerinin sınıflandırılması problemi için matematiksel programlama temelli yeni yöntemler geliştirilmiş ve İstanbul Menkul Kıymetler Borsasında işlem gören şirketlerin derecelendirilmesi probleminin çözümü için uygulanmıştır. Sınıflandırma problemlerinin çözümü için literatürde yer alan geleneksel algoritmalarda şimdiye kadar sıklıkla kullanılan doğrusal fonksiyonlar yerine, projede sunulan bu yöntemlerde, yazarlar tarafından geliştirilen konik fonksiyonlar ve onların çeşitleri kullanılmıştır. Böylece, bu projede, dışbükeylik koşulları gerektirmeyen durumlarda bile yüksek performanslı ayırma kabiliyetine sahip yüzeyler oluşturduğu literatürde de yaygın kabul gören, konik fonksiyonlara dayalı ikili sınıflandırma algoritmaları geliştirilmiştir. Bu algoritmaların çok sınıflı problemlerin çözüm yaklaşımlarında kullanılması, yeni çok sınıflı sınıflandırma yöntemlerinin ortaya çıkmasına olanak tanımaktadır. Hem iki sınıflı, hem de çok sınıflı sınıflandırma problemleri için geliştirilmiş olan bu algoritmalar, bütün dünyada iyi bilinen test problemleri üzerinde denenmiş ve yüksek başarı oranına sahip oldukları gösterilmiştir. Bu sınıflandırıcılar, ilk defa bu projede tanımlanan şirketlerin derecelendirilmesi problemine de uygulanmış ve beklenenlerin üzerinde bir başarı oranı elde edilmiştir. Ayrıca, İMKB'de işlem gören şirketlerin belli bir dönem için çeşitli finansal özellikleri eğitim kümesi şeklinde kullanılarak gelecek bir dönem için bu şirketlerin durumunu tahmin etmeye imkân veren esnek bir karar destek sistemi oluşturulmuştur. Şirketlerin özellik vektörleri kümesi oluşturulurken verilerinin hangi yıldan başlayarak kaç yıllık dönem için alınacağı karar verici (ve ya kullanıcı) tarafından seçiliyor. Bu karar destek sistemi, gelecek dönemin sınırlarının ve şirketlerin kaç sınıfa sınıflandırılacağının da karar verici tarafından belirleneceği şekilde tasarlanmıştır. Örneğin, eğer karar verici, şirketleri sadece kar edenler ve zarar edenler şeklinde sınıflandırmak istiyorsa, iki sınıflı bir sınıflandırma problemi çözülüyor. Bunun yanı sıra kullanıcı, kar ve zarar oranlarını kendisi belirleyerek ikiden fazla sayıda sınıfı tanımlayabilir ve şirketlerin bu sınıflara sınıflandırılacağı çok sınıflı problemin çözülmesini de isteyebilir. Bu durumda sistem, talep edilen sayıda sınıftan oluşan sınıflandırma problemini çözerek karar vericiye, onun tarafından seçilen şirketlerin yine de onun tarafından belirlenmiş olan dönemde hangi sınıfta yer alacağını belli bir başarı oranıyla tahmin edebilmektedir.

Anahtar Kelimeler: Veri madenciliği, iki-sınıflı sınıflandırma, çok sınıflı sınıflandırma, sınıflandırma algoritması, konik fonksiyonlar, tamsayılı programlama, matematiksel programlama, global optimizasyon, çok amaçlı optimizasyon, karar destek sistemi, finansal oranlar, şirketlerin derecelendirilmesi.

Abstract. In this project, new mathematical programming based algorithms for solving two-class and multi-class real life classification problems are developed. These algorithms are applied for solving the company rating problem developed for companies from Istanbul Stock Exchange. One of the main features of the methods developed in this project is that, they use different kinds of conic functions instead of linear ones commonly used in the literature. These conic functions have been developed by the authors and high separability performance of surfaces obtained by these functions, has been demonstrated in publications on this topic even for problems without convexity conditions. Since multi-class classification algorithms generally are based on two-class classification algorithms, methods developed in this project for two-class classification, have a potential for developing different multi-class classification algorithms. The high performances of the new algorithms suggested for both two- and multi-class classification problems, have been demonstrated on well-known benchmark test problems. The classifiers so developed, are applied for solving the company rating problem firstly defined in this project. There was obtained test accuracy over the expected level. Moreover, the flexible decision support system using some subset (selected by decision maker or user) of companies from Istanbul Stock Exchange as a training set, two or multi-class (the number of classes is also selected by a decision maker) classifier is constructed. This classifier uses the set of vectors consisting of different financial ratios or other performance indicators for selected companies, beginning with some year and taking into account the data for several years backward, and then solves the classification problem and determines the classes for the selected companies for some time period ahead of the current year. The number of years and the number of classes are all selected by a decision maker in the decision support system developed.

Keywords: Data mining, two-class classification, multi-class classification, classification algorithm, conic functions, integer programming, mathematical programming, global optimization, multi objective optimization, decision support system, financial ratios, company rating.

1 GİRİŞ

Günümüzde bilgisayar teknolojisinin yüksek hızla gelişmesi sonucu çeşitli alanlarda faaliyetini sürdüren iktisadi, ticari, finansal, tıbbi, endüstriyel ve diğer sistemlerde çok sayıda verinin sistemli bir şekilde elektronik ortamlarda depolanması olanakları ortaya çıkmıştır. Milyonlarca, hatta milyarlarca veri içerisinden belli amaca hizmet edecek verilerin seçilmesi, "ayıklanması" ve bu verilerin, ait oldukları sistemlerin faaliyetini daha etkin kılmak amacıyla kullanılmasında ortaya çıkan problemleri, veri madenciliği problemleri olarak nitelendirebiliriz.

Sınıflandırma problemleri, veri madenciliğinde karşılaşılan problemler içerisinde önemli bir yere sahiptir. Denetimli ve denetimsiz olarak nitelendirilebilen bu türlü problemlerden denetimli veri sınıflandırma problemleri, yönetim bilimi, finans, tıp ve kimya alanlarında karşımıza çıkmaktadır. Denetimli veri sınıflandırmanın amacı, hangi sınıfa ait olduğu bilinen gözlemler eğitim amaçlı kullanılarak, ortaya çıkacak yeni örneklerin sınıflandırılabilmesi için araçlar oluşturmaktır. Sınıflandırma problemleri, ikili ve çoklu sınıflandırma problemleri olarak karşımıza çıkmaktadır.

Tıpta yoğun olarak karşılaşılan benzer problemlerden bir örnek üzerinde sınıflandırma problemini tanımlayalım. Kardiyoloji polikliniğine kalp ağrısı şikâyetleri ile gelen insanlara bir takım testler uygulanır ve bazı tahliller yapılır. Bu işlemlerin sonuçları teşhis için yeterli gelmezse bu insanlara ileri işlemler (örneğin, anjiyo adı verilen ve damar tıkanıklığının derecesini belirlemeye imkân veren bir ameliyat) uygulanır. Bu ileri işlemler hem hastalar için hem de doktorlar için eziyetli ve masraflı olabiliyor. Bu problemi, bir sınıflandırma problemi olarak aşağıdaki şekilde ele alabiliriz. Hastane kayıtlarında var olan ve belli sayıda insana uygulanmış olan aynı test ve tahlillerin rakamsal sonuçlarından oluşan vektörler bizim için veri kümesini oluşturmaktadır. Bu insanlara uygulanmış olan ileri işlemlerin (örneğin, anjiyo ameliyatı) sonucunda bunlardan (hastaneye müracaat den insanlardan) hangisinin hasta, hangisinin hasta olmadığı bilgisi de elimizdedir. Bu bilgileri kullanarak hastalıklı ve sağlıklı insanlar olmak üzere iki sınıf oluşturabiliriz. Problem, bu iki sınıfta yer alan verileri, eğitim kümesi şeklinde kullanarak bir "tanıma" aracı oluşturmaktır, öyle ki bu araç, eğitim kümesinde yer almayan her hangi bir insana ait benzer test ve tahlil sonuçlarından oluşan bir vektörün hastalıklı veya sağlıklı bir insana ait olduğunu belli bir olasılıkla söyleyebilsin. Bu araca sınıflandırıcı, olasılığa ise bu sınıflandırıcının başarı oranı denir. Burada, iki sınıftan oluşan bir problem ele alındığı için, probleme de iki sınıflı veya ikili sınıflandırma problemi denir. Eğer polikliniğe gelen insanlar sadece hastalıklı ve sağlıklı değil de, hastalığın belli bir aşamasında olan insanlar diye "sınıflandırılsalardı" bu durumda çok sınıflı veya çoklu sınıflandırma problemi ile karşı karşıya kalmış olurduk.

Altmışlı yıllardan günümüze kadar veri sınıflandırma probleminin çözümü için önerilmiş algoritmalar genelde matematiksel programlama, istatistiksel, makine öğrenmesi (machine learning) ve sinir ağları (neural networks) yaklaşımlarına dayanmaktadır (bakınız, Fukunaga,1990; Quinlan, 1993; Michie v.d., 1994).

Matematiksel programlama temelli teknikler, kesin çözüm yöntemleri kullanması ve kesin çözüm bulmayı hedeflemesi sebebiyle yukarıdaki yaklaşımlar içerisinde önemli bir yer tutmaktadır. Matematiksel programlama tekniklerinin, denetimli veri sınıflandırma problemlerinin çözümüne uygulanmasında iki ana yaklaşım vardır. Bunlardan birincisi dışsal yaklaşım olup, sınıflandırıcının belirli bir fonksiyon şeklinde tanımlanması temeline dayanmaktadır. Eğitim kümesinde yer alan kümelerin sınıflandırılması, başka bir deyimle ayrılması, farklı kümelerin elemanlarının bu fonksiyona farklı değerler sağlaması şeklinde tanımlanmaktadır. İçsel yaklaşım olarak bilinen ikinci yaklaşımda ise verilen eğitim kümeleri öbek (cluster) merkezlerine yaklaştırılır. Yeni veri vektörü en yakın öbeğe ve bu öbeği içeren sınıfa atanır.

Bu projede, iki ve daha fazla sınıfa sahip gerçek hayat veri kümelerinin ayrılması problemi için dışsal yaklaşıma dayalı matematiksel programlama temelli yeni yöntemler geliştirilmiş ve İstanbul Menkul Kıymetler Borsasında (IMKB) işlem gören şirketlerin derecelendirilmesi probleminin çözümü için uygulanmıştır.

Projede ele alınacak problem, ikiden fazla sınıfa ait nokta (vektör) örnekleri verildiğinde bu sınıfları birbirinden ayıran (farklandıran) ve ileride karşılaşılabilecek benzer örneklerin hangi sınıfa ait olduğunu belli bir başarı oranıyla belirleyebilen ve sınıflandırıcı olarak adlandırılan bir fonksiyonun (veya fonksiyonların) oluşturulmasıdır.

Çok sınıflı sınıflandırma yöntemlerinin temelinde ikili sınıflandırma yöntemleri yer almaktadır. Tıptan ekonomiye, robot teknolojilerinden biyoinformatiğe kadar ve daha birçok alanda yoğun uygulanması nedeni ile ikili sınıflandırma yöntemlerinin kendileri gibi, bunların çok sınıflı problemlere etkin bir şekilde genişletilmesi yöntemleri de günümüzde, bütün dünyada üzerinde yoğun şekilde çalışılan güncel bir araştırma konusudur.

İkili sınıflandırma problemi, R^n uzayında verilmiş, her biri sonlu sayıda noktadan oluşan iki kümenin bir birinden ayrılması problemidir. Bu kümelerin dışbükey örtüleri ayrık olduğu durumda, onları kesin ayıran bir hiperdüzlemin bulunabileceği iyi bilinmektedir. Kümelerin dışbükey örtülerinin kesiştiği durumda belli bir hatalı sınıflandırma oranını en küçükleyecek hiperdüzlemin bulunması için doğrusal programlama tekniği uygulanabilir. Böyle bir hiperdüzlemin bulunması için bir teknik Bennett ve Mangasarian (1992) tarafından tarif edilmiştir. Benzer yaklaşım temelinde çeşitli algoritmalar Chen ve Mangasarian (1995), Bennett ve Blue (1997), Bennett ve Bradley (1997), Bradley v.d. (1999) tarafından da önerilmiştir. Megiddo (1988) polihedral ayrılabilirliği incelemiş ve bu problemin çeşitli versiyonlarının NP-tam olduğunu ispat etmiştir. Astorino ve Gaudioso (2002), Bennett ve Mangasarian (1992) tarafından geliştirilen teknikten yararlanarak doğrusal programlama temelli *h*-polihedral ayırma algoritmasını geliştirmişlerdir. Bagirov (2005), *h*-polihedral ayırmanın bir genelleştirilmesi olarak *max-min* ayrılabilirlik kavramını sunmuş ve ikili sınıflandırma problemlerinin çözümü için etkin bir algoritma geliştirmiştir.

Gasimov ve Ozturk (2006), polihedral fonksiyonların özel bir türünü kullanarak ikili sınıflandırma probleminin çözümü için doğrusal alt-problemlerinin ardışık çözümüne dayanan ve sonlu adımda eğitim kümesinde %100 başarı sağlayan Polihedral Konik Fonksiyonlar (PKF) Algoritmasını önermişlerdir. Algoritmanın her ardıştırmasında kümelerden birinin belli bir kısmını diğer kümenin tamamından ayıran ve grafiği polihedral koni olan bir fonksiyon oluşturulur. Nihai ayırıcı fonksiyon bu fonksiyonların noktasal en-küçüğü olarak tanımlanır. Bilgimiz dâhilinde, literatürde açık şeklide tanımlanmış tek bir ayırıcı fonksiyonun elde edildiği ve bu fonksiyon ile eğitim kümesinde %100 başarı sağlayan başka bir yöntem bulunmamaktadır.

Bu proje kapsamında yapılan çalışmalarda polihedral konik fonksiyonlar sınıfı farklı normlar kullanılarak genişletilmiş ve bu fonksiyonların kullanılmasıyla PKF algoritması geliştirilerek çeşitlendirilmiştir. Geliştirilen algoritmaların sonlu adımda tamamlandığı ve bu algoritmalar kullanılarak herhangi iki ayrık, sonlu sayıda noktadan oluşan kümenin tam olarak ayrılabilirliği gösterilmiştir. Bunun dışında genelleştirme özelliğine sahip tamamen yeni bir matematiksel modelin çözümlerine dayalı bir ikili sınıflandırma yöntemi önerilmiştir. İyi bilinen, gerçek hayat veri kümelerini içeren sınıflandırma problemleri üzerinde gerçekleştirilen uygulamalar, önerilen ikili sınıflandırma yöntemlerinin doğruluk ve çalışma zamanı bakış açılarından yüksek performans sergilediğini göstermiştir. Bu ikili sınıflandırma algoritmaları farklı yaklaşımlara dayanan çoklu sınıflandırma yöntemlerinin geliştirilmesinde kullanılmıştır. Yapılan bu çalışmalar tamamen orijinal olup, aslında hem ikili hem de çoklu sınıflandırma problemlerinin çözümü için çok sayıda yeni yöntemin ortaya çıkmasına öncülük edecek niteliktedir. Bir proje kapsamında bütün bunların denenmesi ve uygulanması mümkün olmamakla beraber gelecek çalışmalara ışık tutacak birçok test de başarıyla tamamlanmıştır. Geliştirilen bu yöntemlerin bir kısmı bu proje kapsamında tarif edilen orijinal bir problem olan İMKB'de işlem gören şirketlerin getiri oranlarına göre derecelendirilmesi problemine uygulanmıştır.

Bu raporun ikinci bölümünde, ele alınan problem detaylı bir şekilde tanımlanmıştır. Bu bölümde ikili ve çoklu sınıflandırma yöntemleri ile ilgili bilgiler verilmiş ve IMKB'de işlem gören şirketlerin derecelendirmesi problemi için taban teşkil edecek finansal bilgiler ayrıntılı bir şekilde açıklanmıştır. Üçüncü bölümde yeni geliştirilen ikili ve çoklu sınıflandırma algoritmaları açıklanmış ve test problemleri üzerinde denemeler yapılmıştır. Dördüncü bölümde algoritmaların uygulanması için geliştirilen sistem tanıtılmış ve proje kapsamında tanımlanan problem için yapılan denemeler ile ilgili hesapsal sonuçlar verilmiştir. Son bölümde proje kapsamında yapılanlar ile ilgili sonuçlar yorumlanmış ve gelecek çalışmalar ile ilgili ipuçları verilmiştir.

2 PROBLEMİN TANIMLANMASI

Bu projede, İstanbul Menkul Kıymetler Borsasında (IMKB) işlem gören şirketlerin derecelendirilmesi problemi çok sınıflı sınıflandırma problemi olarak aşağıdaki şekilde tanımlanmıştır.

İMKB'de işlem gören şirketlere ait 1998-2007 yıllarını kapsayan 3 aylık, 6 aylık, 9 aylık ve yıllık veriler alınmıştır. Bu veriler, şirketlere ait getiri, kar oranları, çeşitli bilançolar, finansal göstergeler şeklindedir. Bu veriler içerisinden aynı dönemler için (örneğin 2002-2006 yıllarında yıl bazında) aynı göstergelere sahip şirketlerin listesinin oluşturulduğunu varsayalım. Bu şirketlerin hepsi için ilgili dönemlerde şirketlerin finansal durumunu nitelendirecek çeşitli finansal oranlar hesaplanarak bu oranlardan oluşan vektörlerin kümesi elde edilir. 2002-2006 yılları için verileri oluşturulmuş olan bu şirketlerin 2007 yılındaki performanslarına göre (örneğin kar düzeyine göre) sınıflar oluşturulur. Projede, bu sınıflardan oluşan kümeler eğitim kümesi olarak kullanılmış ve bir sınıflandırma yöntemi uygulanarak bir sınıflandırıcı oluşturulmuştur. Bu sınıflandırıcının başarı oranı, oluşturulacak test kümelerinin yanı sıra, 2007 yılında yaşanmış olan gerçek veriler üzerinde de denenmiştir. Sonuçta, geçmiş yıllardaki performansları göz önünde bulundurularak gelecek yıl (veya gelecek farklı dönemler için) şirketleri derecelendirmek (bir başka deyimle, kullanıcıyı ilgilendirecek dönemde ilgili şirketlerin finansal durumunu tahmin etmek) mümkün olabilmiştir.

Bu problemin çözümü için günümüzde var olan çeşitli çözüm yaklaşımlarının yanı sıra ilk kez bu projede önerilecek yeni yaklaşımlar da uygulanmıştır. Projede, farklı algoritmalar kullanılarak farklı sınıflandırıcılar oluşturulmuş ve bu algoritmaların performansları başarı oranlarına göre kıyaslanmıştır.

2.1 Sınıflandırma Yöntemleri

Çok-sınıflı problemlerin çözümü için sınıflandırıcıların oluşturulması yöntemleri genellikle ikili sınıflandırma yöntemlerine dayanmaktadır. Tarihen, çok-sınıflı sınıflandırıcıların oluşturulması için, destek vektör makineleri (support vector machines), sinir ağları ve matematiksel programlama teknikleri kullanılmıştır. İkili sınıflandırma yöntemlerine dayanan üç temel yaklaşım vardır: “bire-karşı-hepsi”, “bire-karşı-bir” ve “yönlü çevrimsiz serim destek vektör makineleri” (directed acyclic graph Support Vector Machines).

Polihedral konik fonksiyonlar (PKF) algoritması, doğrusal fonksiyonların l_1 normu ile genişletilmesi ile ortaya çıkan konik fonksiyonları kullanan etkin bir algoritmadır. Yukarıda, bu algoritmanın eğitim kümesinde sonlu adımda 100% başarı sağladığı vurgulanmıştı. Bu algoritmanın en cazip yönlerinden biri, her adımda doğrusal bir problemin çözülmesi ile kısmi sınıflandırıcının elde edilmesidir. Bir diğer önemli özelliği de kullandığı parametre sayısının, sınıflandırılması istenen kümelerin bulunduğu uzay boyutunun sadece bir fazlasına eşit olmasıdır. Hatırlatalım ki, doğrusal ayırma yöntemleri bu boyut sayısına eşit miktarda parametre kullanıyor ve karmaşık kümelerden (yani, doğrusal olarak ayrılamayan kümeler durumunda) oluşan durumlarda etkili sayılacak bir sınıflandırıcının oluşturulması için hiçbir ümit vermiyor. PKF algoritmasında kullanılan fonksiyonlar, doğrusal fonksiyona l_1 normunun katları eklenerek oluşturulmuştur.

Bu projede, doğrusal fonksiyonlara farklı normların eklenmesiyle yeni fonksiyon sınıfları oluşturulmuş ve ikili sınıflandırma problemleri için PKF benzeri yeni ve farklı performansa sahip algoritmalar ortaya konulmuştur.

Sözü edilen bu yeni yaklaşımları kullanarak, çoklu sınıflandırma problemleri için de yeni yöntemler önerilmiştir.

2.1.1 İkili sınıflandırma yöntemleri

R^n'de, sonlu sayıda noktaya sahip A ve B gibi iki ayrık kümenin ayrılması problemi için matematiksel programlamanın ilk kullanımları altmışlı yılların sonlarına rastlamaktadır.

2.1.1.1 Doğrusal ayırma

Bennett ve Mangasarian (1992) tarafından yapılan çalışmada, ayırıcı hiperdüzlemin iki kümeye de belli bir mesafede olmasını sağlayacak şekilde doğrusal ayrılabilirlik aşağıdaki gibi tanımlanmıştır.

A ve *B* sırasıyla *m* ve *p* adet her biri *n* boyutlu vektörden oluşan kümeler olsun.

$$
\begin{aligned}
A &= \{a^1,\ldots,a^m\}, \quad a_i \in R^n, \qquad i=\{1,\ldots,m\}, \\
B &= \{b^1,\ldots,b^p\}, \quad b_j \in R^n, \qquad j=\{1,\ldots,p\}
\end{aligned}
$$

Bu kümeleri ayıracak olan hiperdüzlemi bulmak için

$$
f(x,y) = \frac{1}{m}\sum_{i=1}^{m}\max\left(0,\left\langle x,a^i\right\rangle - y + 1\right) + \frac{1}{p}\sum_{j=1}^{p}\max\left(0,\left\langle x,b^j\right\rangle + y + 1\right)
$$

olmak üzere

$$\min f(x,y)$$

problemi çözülmektedir. Burada $\langle \cdot,\cdot \rangle$ gösterimi *n* boyutlu uzayda skaler çarpıma karşı gelmektedir. Bu problemin aşağıda verilen doğrusal programlama problemine eşdeğer olduğu gösterilmiştir.

$$t_i \geq \left\langle x, a^i \right\rangle - y + 1, \qquad i = 1, \ldots, m$$

$$z_j \geq -\left\langle x, b^j \right\rangle + y + 1, \qquad j = 1, \ldots, p$$

$$t, z \geq 0$$

kısıtları altında

$$\min \quad \frac{1}{m}\sum_{i=1}^{m} t_i + \frac{1}{p}\sum_{j=1}^{p} z_j$$

A ve *B* kümeleri ancak ve ancak bu problemin amaç fonksiyonu sıfıra eşit olduğunda doğrusal programlama ile tam olarak ayrılabilir.

2.1.1.2 h-polihedral ayırma

Astorino ve Gaudioso (2002), *h* adet hiperdüzlem ile bir dışbükey polihedron üreterek *n*-boyutlu uzayda sonlu sayıda noktadan oluşan iki *A* ve *B* kümesi için ayırma problemini incelemişlerdir. Eğer *B* ile *A*'nın dışbükey örtüsünün kesişimi boş küme ise bu iki küme (polihedral ayrılabilirlik) tam olarak ayrılabilir. Kısaca *h*-polihedral ayırma aşağıdaki gibi tanımlanmıştır.

A ve *B* kümeleri, eğer aşağıdaki koşulları sağlayan *h* adet $\{x^i, y^i\}$, $x^i \in R^n$, $y^i \in R$, i=1,…,h, çiftinin oluşturduğu hiperdüzlemler kümesi mevcut ise *h*-polihedral olarak ayrılabilir.

1. herhangi bir *j* =1,…,*m* ve *i*=1,…*h* için

$$\left\langle x^i, a^j \right\rangle - y_i < 0.$$

2. herhangi bir *k* =1,…,*p* ve en az bir $i \in \{1,\ldots h\}$ için

$$\left\langle x^i, b^k \right\rangle - y_i > 0$$

olmalıdır.

A ve *B* kümelerinin

$$h \leq |B|$$

olmak üzere, ancak ve ancak

$$conv(A) \cap B = \phi$$

şartı sağlanırsa *h*-polihedral ayırma ile tam olarak ayrılabileceği Astorino ve Gaudioso (2002) tarafından gösterilmiştir.

Burada, hata fonksiyonu olarak adlandırılan $f(x,y)$ fonksiyonu

$$f(x,y)=\frac{1}{m}\sum_{i=1}^{m}\max\left[0,\max_{1\le i\le h}\left\{\left\langle x^i,a^j\right\rangle-y_i+1\right\}\right]+\frac{1}{p}\sum_{k=1}^{p}\max\left[0,\min_{1\le i\le h}\left\{-\left\langle x^i,b^k\right\rangle+y_i+1\right\}\right]$$

gibi tanımlandığında, *A* ve *B* kümelerinin *h*-polihedral ayrılabilirliği izleyen probleme indirgenmiştir:

$$\min f(x,y)\quad,\quad (x,y)\in R^{(n+1)\times h}$$

Astorino ve Gaudioso tarafından tanımlanan hata fonksiyonu parçalı doğrusal yapıdadır. *A* ve *B* kümeleri ancak ve ancak, $\left(\bar{x}^i,\bar{y}\right)$ problemin bütünsel (global) en iyi çözümü iken,

$$f(\bar{x},\bar{y})=0$$

İse, *h*-polihedral anlamda tam olarak ayrılabilir. Araştırmacılar tarafından bu problemin çözümü için doğrusal programlama temelli bir algoritma önerilmiştir.

2.1.1.3 max-min ayırma

Bagirov (2005) tarafından geliştirilen *max-min* ayırma yaklaşımında yine, sayıları önceden belirlenmiş olan hiperdüzlemler kullanılmakta, ancak bu hiperdüzlemler için farklı alt kümeler tanımlanmaktadır. Bu sayede dışbükey olmayan ayırıcı yüzeyler oluşturulabilmektedir.

Max-min ayırma,

$$x=\{x^1,\ldots,x^l\}\in R^{l\times n}\quad,\quad y=\{y_1,\ldots,y_l\}\in R^l$$

olduğunda ve hata fonksiyonu $f(x,y)$ aşağıdaki gibi tanımlandığında

$$f(x,y)=\frac{1}{m}\sum_{k=1}^{m}\max\left[0,\max_{i\in I}\min_{j\in J_i}\left\{\left\langle x^j,a^k\right\rangle-y_j+1\right\}\right]+\frac{1}{p}\sum_{t=1}^{p}\max\left[0,\min_{i\in I}\max_{j\in J_i}\left\{-\left\langle x^j,b^t\right\rangle+y_j+1\right\}\right]$$

izleyen problem ile ifade edilmektedir.

$$\min f(x,y)\quad,\quad (x,y)\in R^{(n+1)\times l}$$

Bagirov (2005) max-min ayırmanın özel durumda hem *h*-polihedral hem de doğrusal ayırma yaklaşımlarını içerdiğini göstermiştir.

2.1.1.4 Polihedral Konik Fonsiyonlar (PKF) algoritması

Gasımov ve Öztürk, ayırma problemlerinde kullanılmak üzere polihedral fonksiyonların özel bir sınıfı olan polihedral konik fonksiyonları tanımlamışlardır.

Grafiği koni ve seviye kümesi dışbükey polihedron olan bu fonksiyonlar izleyen şekilde tanımlanmıştır:

$$g_{(w,\xi,\gamma)}(x) = w'(x-a) + \xi\|x-a\|_1 - \gamma$$

Burada

$$w \in R^n,\ \xi \in R^+ = [0,+\infty),\ \gamma \geq 1,\ w'x = w_1x_1 + ... + w_nx_n$$

ve

$$\|x\|_1 = |x_1| + \ldots + |x_n|$$

ise *x* vektörünün l_1 normudur.

Bu fonksiyonları kullanarak R^n de verilen *A* ve *B* gibi iki sonlu nokta kümesini ayıran algoritma izleyen adımlardan oluşmaktadır.

PKF algoritması

A ve *B* kümeleri R^n'de verilmiş iki küme olsun:

$$A = \{a^i \in R^n : i \in I\}, \qquad I = \{1,\ldots,m\}$$
$$B = \{b^j \in R^n : j \in J\}, \qquad J = \{1,\ldots,p\}$$

PKF algoritması izleyen şekilde ifade edilir.

Başangıç Adım :

$$l = 1,\ I_l = I,\ A_l = A$$

atamalarını yap, *Adım1*'e git.

Adım 1 : a^l noktası A_l kümesinin keyfi bir noktası olsun. Aşağıda verilen P_l problemini,

$$w \in R^n,\ \xi \in R^+ = [0,+\infty),\ \gamma \geq 1, y \geq 0$$

karar değişkenlerine göre çöz:

$$w'(a^i - a^l) + \xi\|a^i - a^l\|_1 - \gamma + 1 \leq y, \qquad \forall i \in I_l$$
$$-w'(b^j - a^l) - \xi\|b^j - a^l\|_1 + \gamma + 1 \leq 0, \qquad \forall j \in J$$

Kısıtları altında

$$\min\left(\frac{y'e_m}{m}\right) \qquad (P_l)$$

P_l probleminin bir çözümü

$$w^l, \xi^l, \gamma^l, y^l \text{ ve } g_l(x) = g_{(w^l,\xi^l,\gamma^l,y^l)}(x)$$

olsun. *Adım2*' ye git.

Adım 2 :

$$I_{l+1} = \{i \in I_l : g_l(a^i) > 0\},\ A_{l+1} = \{a^i \in A_l : i \in I_{l+1}\},$$

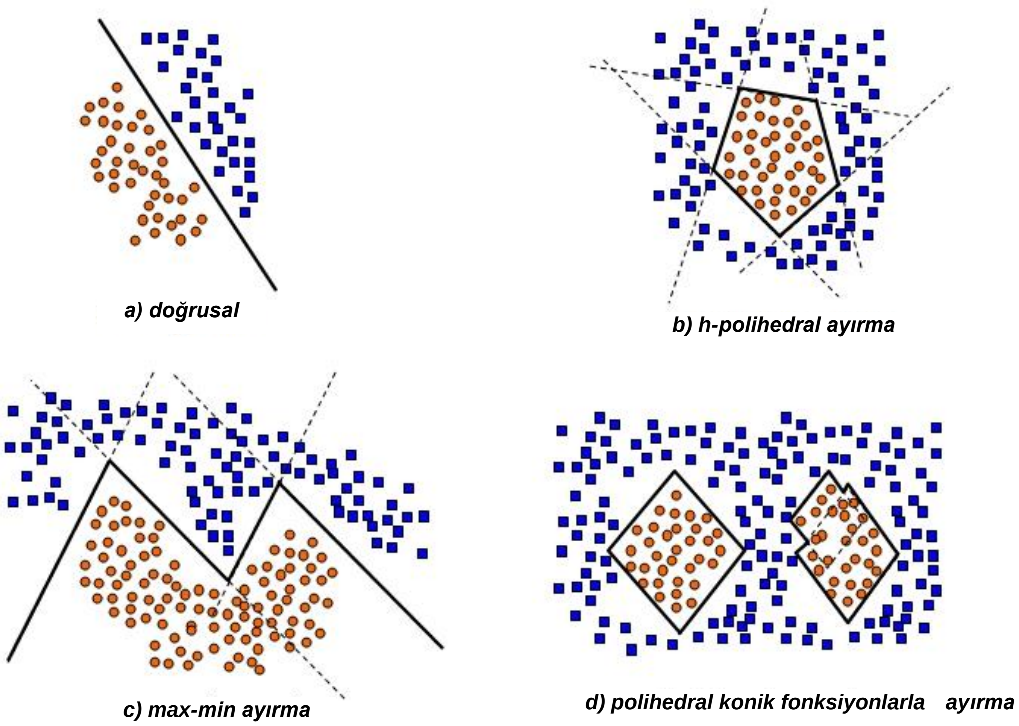
l=*l*+1 olsun. Eğer $A_l \neq \emptyset$ ise *Adım1*'e git.

Adım 3 : *A* ve *B* kümelerini ayıran *g*(*x*) fonksiyonunu aşağıdaki gibi tanımla ve dur.

$$g(x) = \min_l g_l(x).$$

Polihedral konik fonksiyonlar sayesinde bu algoritmanın her ardıştırmasında *A* kümesinden en az bir noktayı *B* kümesinden ayırdığı ispatlanmıştır. Böylece sonlu adımda algoritma *A* kümesini *B*'den ayıracak olan ayırıcı fonksiyonu oluşturuyor.

Literatürdeki matematiksel programlama temelli ikili ayırma yaklaşımlarının R^2'deki geometrik yorumları Şekil 2.1'de verilmiştir. Bu şekilde *A* kümesi yuvarlak noktalar *B* kümesi ise kare noktalar ile temsil edilmektedir.



Şekil 2.1 Matematiksel programlama temelli ikili ayırma yaklaşımları (Gasımov ve Öztürk, 2006)

Doğrusal ayırma ile, dışbükey örtüleri kesişmeyen *A* ve *B* kümelerinin tam olarak ayrıldığı (a), *h*-polihedral ayırma ile *A*'nın dışbükey örtüsü ile *B*'nin kesişiminin boş olduğu *A* ve *B* kümelerinin tam olarak ayrıldığı (b), *max-min* ayırma ile dışbükey olmayan bir yapı sayesinde *A* ve *B* kümelerinin tam olarak ayrıldığı (c) ve son olarak polihedral konik ayırma ile herhangi iki *A* ve *B* kümesinin tam olarak ayrıldığı (d) şekilde gösterilmiştir. Bu yöntemlerden doğrusal ayırmanın (b), (c) ve (d) durumlarında, h-polihedral ayırmanın (c) ve (d) durumlarında ve *max-min* ayırmanın (d) durumunda *A* ve *B* kümelerini tam olarak ayıramayacağı gözükmektedir.

.

Tablo 2.1 İki sınıflı test veri kümeleri üzerinde elde edilmiş on kez çapraz doğrulama sonuçları (Gasımov ve Ozturk, 2006)

Veri Kümesi	Eğitim	Test	Süre (sn)	Çözülen Ortalama DP sayısı
1 Liver	100	68.4	16.84	114
2 WBCD	100	100	0.58	1
3 WBCP	100	75.77	4.44	18
4 Ionosphere	100	88.03	0.84	5
5 Heart	100	79.12	10.23	74
6 Diabetes	100	71.48	44.17	216

Gasımov ve Ozturk (2006), literatürde iyi bilenen iki sınıflı test problemlerini *PKF* algoritması ile araştırmış ve elde edilen sonuçları Tablo 2.1'deki gibi sunmuşlardır. Tablodan da görüldüğü gibi tüm problemlerde eğitim başarısı %100 olmuş ve test başarıları literatür ile rekabet edebilecek düzeyde gerçekleşmiştir. Özellikle WBCD (Winsconsin Breast Cancer Diagnosis) "Winsconsin Göğüs Kanseri Teşhisi" veri kümesinde, tek bir PKF ile test aşamasında %100 başarı sağlanmıştır. Bu da *PKF*'lerin ayırmadaki başarısını ortaya koymaktadır.

Projede, ikili sınıflandırma algoritması olarak *max-min* ve *PKF* algoritmalarının yanı sıra, PKF algoritmasında kullanılan

$$g_{(w,\xi,\gamma)} = w'(x-a) + \xi\|x-a\|_1 - \gamma$$

fonksiyonunun yerine

$$g_{(w,\xi,\sigma,\gamma)} = w'(x-a) + \xi\|x-a\|_1 + \sigma\|x-a\|_2 - \gamma$$

ve/veya

$$g_{(w,\sigma,\gamma)} = w'(x-a) + \sigma\|x-a\|_2 - \gamma$$

şeklinde farklı normlar içeren fonksiyonlar da kullanılmıştır. Bu şekilde ayırıcı fonksiyonlar da PKF ler gibi literatürde daha önce kullanılmamış olup ilk kez bu projede önerilmektedir.

PKF algoritmasında yeni tip ayırıcı fonksiyonların kullanılmasının yanı sıra, bu projede tamamen farklı yeni bir algoritma daha önerilmiştir. Bu yeni algoritmada da yukarıda sözü edilen her iki sınıf ayırıcı fonksiyonlar kullanılmıştır. Yeni algoritmanın PKF'den farkı, her iterasyonda A kümesinin B'den ayrılmış elemanlarının dışarıda bırakılması yerine, A'nın bütün elemanlarını tek tek, oluşturulacak ayırıcı polihedral koninin merkezi olarak alıp, en az *A* kümesinin elemanları sayıda ayırıcı fonksiyonlar oluşturulması şeklinde düşünülmektedir. Daha sonra nihai sınıflandırıcı, tüm ayırıcı fonksiyonların noktasal minimumu olarak tanımlanır. Yeni algoritmanın oluşturacağı ayırıcı fonksiyon sayısı, PKF'ninkine göre daha fazla olabilir, fakat bilgisayar programı daha sade olması nedeniyle, hesaplama karmaşıklığı açısından daha komplike değildir.

2.1.2 Çoklu sınıflandırma yöntemleri

Çok sınıflı problemler R^n'de, her biri m_i sayıda noktadan oluşan k tane A_i, i=1,...,k, kümelerinin ayrılması problemidir. Bu problemler literatürde çoğunlukla destek vektör makineleri ve sinir ağları kullanılarak incelenmiştir. Bunun yanında matematiksel programlama (Bredensteiner ve Bennett, 1999; Bagirov ve Ugon, 2005) bakış açısı ile de bu problem ele alınmıştır. Çoklu sınıflandırma problemlerinin çözümü için, temelinde ikili sınıflandırma yöntemlerinin kullanıldığı bire-karşı-hepsi (1-e-h) (one-against-all), bire-karşı-bir (1-e-1) (one-against-one) ve yönlü çevrimsiz serim destek vektör makineleri (directed acyclic graph SVM) gibi yaklaşımlar önerilmiştir.

Bagirov ve Ugon (2005), bire-karşı-hepsi yaklaşımını kullanarak *max-min* ayırma (Bagirov, 2005) yöntemini çok sınıflı problemleri çözmek üzere genişletmişlerdir. Bredensteiner ve Bennett (1999), doğrusal programlama ve destek vektör makineleri yaklaşımlarını birleştirerek yeni yaklaşımlar türetme üzerine çalışmışlardır.

Literatürde görüldüğü gibi çok sınıflı problemler için birçok algoritma geliştirildiği halde herhangi bir yaklaşımın diğerlerine tam olarak üstünlük sağladığı iddia edilememekte ve ikili sınıflandırma yaklaşımlarının etkin bir şekilde çoklu sınıflandırma yaklaşımlarına genişletilmesinin hala güncelliğini koruyan bir araştırma alanı olduğuna dikkat çekilmektedir (Hsu ve Lin, 2002).

Bu projede şimdiye kadar bilinen bire karşı bir ve bire karşı hepsi yaklaşımlarının yanı sıra, tarafımızdan önerilen ardışık sınıflandırma diye adlandırdığımız yeni bir yaklaşım da kullanılacaktır. Bu yaklaşımlar;

- Bire karşı hepsi (1-e-h),
- Bire karşı bir (1-e-1),
- Ardışık sınıflandırma,

olarak ele alınacaktır.

2.2 İlgilenilen dönemlerdeki IMKB'de işlem gören şirket bilançolarının toplanması ve kullanılacak hale getirilmesi

2.2.1 Verilerin toplanması

İstanbul Menkul Kıymetler Borsası (İMKB), hisse senetleri, hazine bonoları ve devlet tahvilleri, gelir ortaklığı sertifikaları, özel sektör tahvilleri, yabancı menkul kıymetler, gayrimenkul sertifikaları ve uluslararası menkul kıymetlerin alım ve satımının yapılmasını sağlamak amacıyla 26 Aralık 1985 günü kurulmuş olup, 3 Ocak 1986 yılından bu yana faaliyet göstermektedir. İMKB'de işlem gören şirketlerle ilgili verilerin yayın hakkı İMKB'ye ait olmakla birlikte 02.07.1998 tarihinden itibaren "İMKB Veri Yayın Sözleşmesi" kapsamında yayın yapan kuruluşlar tarafından da ücretli olarak yayınlanmaya başlanmıştır. İMKB Hisse Senetleri Piyasası Verileri, İMKB'nin internet sayfasından ücretsiz olarak indirilebilmektedir (www.imkb.gov.tr). 2007 yılı sonu itibariyle Ulusal Pazar'da işlem gören Şirket sayısı 292'dir. Proje kapsamında özellikle 1998–2007 yılları arasında İMKB'de işlem gören şirketlerin aylık ortalama alım-satım fiyatları, işlem hacimleri, 3, 6, 9 ve 12 Aylık Mali Tabloları verilerinin elde edilmesi hedeflendi. Bu hedefe ulaşmak için öncelikle İMKB'nin web sayfasından ilgili veriler indirilmeye çalışıldı. Ancak verilerin doğrudan web sayfasından indirilmesi yolu, şirket ve yıl bazında ayrı ayrı indirilmelerini gerektirmesi ve yoğun işlem yüküne sebep olmasından dolayı tercih edilmedi. İMKB, akademik araştırma çalışmalarında kullanılacak verilerin tedarikini ücretsiz olarak gerçekleştirmektedir. TÜBİTAK Projesi kapsamında ihtiyaç duyulan veriler de akademik amaçlı olarak kullanılacağından doğrudan İMKB'den temin edilmesi yoluna gidilmiştir. İMKB Eğitim ve Yayın Müdürlüğü'nden Sayın Özkan Çevik ile yapılan resmi yazışmalar sonucunda 15 adet CD-R ekte verilerek 01.01.1998 – 31.12.2007 tarihleri arasında "İstanbul Menkul Kıymetler Borsası'nda işlem gören şirketlerin Microsoft Excel ortamındaki 3, 6, 9 ve 12 Aylık Mali Tabloları" 29.03.2008 tarihinde ücretsiz olarak temin edilmiştir. 1998 – 2007 yılları arasında İMKB'de faaliyet gösteren şirketlerin mali tablolarından hareketle şirketlerin finansal özellik vektörlerinin oluşturulması için öncelikle mali tablolar ve finansal oranlarla ilgili bilgi, denklem ve literatür araştırılmıştır.

2.2.2 Şirketlerin finansal tabloları ve oranları

Firmaların mali durumları ile faaliyet sonuçlarının değerlendirilmesi, gerek firma yöneticilerine ve finans kuruluşlarına gerekse firma yatırımcılarına kritik kararları alırken çok değerli bilgiler sağlamaktadır (Walter and Robert, 1988). Firmaların performanslarını belirlemek, değerlendirmek ve geçmiş verilerle ya da diğer firmalarla karşılaştırabilmek için kullandıkları tablolara Finansal Tablolar denilmektedir (Ercan ve Ban, 2005). Bilânço ve gelir tablosu temel finansal tablolar olarak adlandırılmaktadır. Finansal tabloların kullanıcılara yararlı ve karşılaştırılabilir olmaları için genel kabul görmüş muhasebe ilkelerine göre hazırlanmaları ve finansal tablolarda yer alan bilgilerin güvenilir olması gerekmektedir. Bilânço, firmanın belirli bir tarihte sahip olduğu varlıklar ile bu varlıkların sağladığı kaynakları gösteren finansal bir tablodur (Ercan ve Ban, 2005). İşletmenin sahip olduğu varlıklar, mal ve hizmet üretimi için ayrılmış iktisadi değerler olup, bu değerler özsermaye ve yabancı kaynaklarla sağlanmış olmaktadır. Bilânço, Aktif ve Pasif olarak adlandırılan iki ana bölümden oluşmaktadır. Çift taraflı hesap sisteminin gereği olarak aktif ve pasif toplamlarının birbirine eşit olması gerekmektedir. Aktif, işletmenin sahip olduğu varlıkları ifade ederken, Pasif de, bu varlıkların finanse edildiği özsermaye ve yabancı kaynakları göstermektedir. İMKB'de işlem gören şirketler 3, 6, 9 ve 12 aylık olmak üzere yılda dört kez bilânço açıklamakta ve bunlar bağımsız denetimden geçmektedirler.

Bir firmanın mali durumu, faaliyet sonuçları kavranırken ve değerlendirilirken, firmanın mali tablolarında görülen rakamlardan çok, bilanço ve gelir tablosunda yer alan kalemler arasındaki ilişkiler daha anlamlı olmakta, bu nedenle finansal analizde, finansal oranlardan geniş ölçüde yararlanılmaktadır (Akgüç, Ö., 1994, Walter and Robert, 1988). Finansal oranlar, mali tablolarda yer alan herhangi iki kalem arasındaki ilişkinin basit matematiksel ifadesi olarak isimlendirilebilir. Mali tablolarda yer alan kalemlerin sayısına göre, geometrik olarak artan sayıda oran hesaplamak, bütün kalemleri birbiriyle karşılaştırmak geniş bir oranlar kümesi oluşturmak mümkündür. Ancak gerek mali analistler, gerekse yöneticiler açısından önemli olan, firmanın likitide durumunu, borç ödeme gücü, finansman şekli, faaliyet sonuçları, karlılığı, iktisadi varlıklarını etkin biçimde kullanıp kullanmadığı konularındaki cevap verecek oranları hesaplamaktır. Analiz açısından önemli olan çok sayıda oran hesaplamak değil; az sayıda fakat anlamlı oranlar hesaplamaktır.

Finansal oranlar, firmanın yalnız geçmiş ve cari mali durumunu değerlendirmek açısından değil, planlama ve kontrol işlevini yerine getirmede ve üçüncü kişilerin özellikle firma ortaklarının, hisse senetlerine yatırım yapmak isteyen birikim sahiplerinin ve finansman kurumlarının, firmanın mali durumunu ve gücünü nasıl gördüklerini, değerlendirdiklerini kestirme yönünden de finans yöneticisine yararlıdır. Finansman kurumları, firmanın kredi değerliliğini saptarken, oranlar analizinden yararlandıkları gibi, birikim sahipleri de yatırım kararlarında firmaların finansal oranlarını dikkate almaktadırlar. Firmanın finansal oranlarının yeterli, doyurucu bulunması, finansman kurumlarından kredi sağlanmasını kolaylaştırdığı gibi, birikim sahipleri açısından da hisse senetlerini çekici hale getirmektedir (Akgüç, Ö., 1994).

Finansal analizde, sadece oranları hesaplamak tek başına yeterli değildir. Finansal oranları hesaplamak analizin mekanik yönüdür. Önemli olan hesaplanan oranların yorumlanması, değerlendirilmesidir.

Finansal oranlarla hisse senedi getirileri arasındaki ilişkiyi inceleyen çalışmaların çoğu, uygun finansal oranlara sahip şirket hisse senetlerinin yüksek getiri potansiyeline sahip olabileceğini söylemektedir. Getiri tahmininde kullanılan önemli finansal oranlar, fiyat/kazanç oranı (F/K), piyasa değeri/defter değeri oranı (PD/DD) ile likidite, karlılık ve sermaye yapısı oranlarıdır. Canbaş v.d.(1997), 1993-1997 döneminde hisse senetleri İ.M.K.B'de işlem gören 173 endüstri işletmesinin finansal oranlarını kullanarak yaptığı çalışmada, finansal oranların hisse senetlerinin değerlemesinde ve hisse senetlerinin getirilerinin açıklanmasında yararlı olduğunu, yatırımcı açısından yararlı bilgi sağlayan oranların da likidite, finansal yapı ve karlılık oranları olduğunu söylemektedir. Bağımsız değişken seti olarak finansal oranların, bağımlı değişken seti olarak da hisse senedi getirilerinin esas alındığı bir diğer çalışmada Demir v.d. (1997) finansal oranların hisse senedi getirisini açıklamada anlamlı sonuçlar ortaya koyduğu, fakat F/K oranı ile hisse senedi getirisi arasında anlamlı bir ilişkinin bulunamadığını, "F/K oranı düşük olan hisse senedinin getirisinin daha yüksek olacağı" kanısının İ.M.K.B. için doğru olmadığı sonucuna ulaşılmıştır. Easton ve Haris (1991), Strong (1993) karlardaki değişmenin hisse senedi getirileriyle yüksek bir ilişki içinde olduğunu göstermişlerdir. Özer, (1996) İ.M.K.B.'de yaptığı araştırmada şirketlerin karlılıklarıyla hisse senedi getirileri arasında tutarlı ilişkiler bulmuştur. İ.M.K.B'de sanayi şirketlerinin performanslarını finansal göstergelerle tahmin eden bir diğer çalışma da (Alkan, 1997) şirketlerin görünümlerinin belirlenmesinde borç göstergelerinin etkili olduğu, vergi öncesi karların, öz sermaye oranının oldukça önemli olduğu ve yatırımcıların şirket değerlendirme kriterlerine bu oranın yansıması gerektiği belirtilmektedir.

Çalışmada geliştirilen modellerde temel özellik vektörü olarak şirketlerin finansal oranları kullanılmaktadır. Teorik olarak hesaplanabilecek oran finansal oran sayısı yüzlerle ifade edilebilir. Bunlar içinde, literatürde üzerinde birleşilen, önemli olduğu kabul edilen, yaygın olarak kullanılan ve kolay hesaplanabilir oranlar seçilmiştir (Akgüç, 1994, Aktaş v.d, 2003, Ercan ve Ban, 2005, Yalçıner v.d., 2005, Walter and Robert, 1988, Wang and Lee, 2008).

2.2.2.1 Likidite Oranları

Likidite oranları, işletmenin likiditesi açısından dönen varlıkların ne oranda güvenli olduğunu gösteren oranlardır. İşletmenin likidite durumunu, vadesi gelen borçları ödeme gücünü; beklenmedik piyasa şartlarında ve ekonomik durumlarda işletmenin faaliyetlerini sürdürebilme yetisini gösteren bu oranlar işletmeye borç verenler açısından çok önemlidir.

Cari Oran (CO)

Cari oran, şirketlerin faaliyetlerini devam ettirebilmeleri için gerekli olan brüt işletme sermayesinin (dönen varlıkların) kısa vadeli borçlara bölünmesi ile hesaplanır.

$$\text{Cari Oranı} = \frac{\text{Dönen Varlıklar}}{\text{Kısa Vadeli Borçlar}}$$

Cari oran bize basit olarak net işletme sermayesinin yeterliliğini ve borç ödeyebilme kapasitesini göstermektedir. Buna rağmen net işletme sermayesi değişimi ile birlikte değerlendirilmesi daha doğru olmaktadır. Net işletme sermayesinin artıyor olması şirketin borç ödeme kapasitesinin arttığının göstergesi değildir. Söz konusu işletmede kısa süreli borçlar dönen varlıklardan hızlı artmışsa cari oran düşebilir. Ayrıca cari oran değerlendirilirken dönen varlıkların kalitesi de (satışa çıkarıldığında gerçek değerini bulması) göz önüne alınmalıdır. Batılı finans kurumlarında bu oranın 2 olması genel kabul gördüğü halde, yüksek enflasyonlu ve kıt fon kaynaklarına sahip ülkelerde endüstri ve sektörlerin değişik özellikleri göz önünde bulundurulduğunda cari oranın 1,5 olması genellikle yeterli olarak kabul edilmektedir (Akgüç, 1994). Sermaye, duran varlıklara yatırım (veya satış) ve uzun vadeli borçlanma hareketleri cari oranda değişikliğe sebep olmaktadır. Bu yüzden, iki dönem arasındaki cari oran değişikliklerini iyi incelemek gerekmektedir.

Likitide Oranı (LO)

Likidite oranı (asit-test oranı), cari oranın geliştirilmiş ve daha anlamlı hale getirilmiş hali olarak düşünülebilinir. Likidite oranında, dönen varlıklar içinde görülen fakat nakite kolayca çevrilemeyen stoklar çıkarılır ve daha sonra duran varlıklar kısa vadeli borçlara bölünür.

$$\text{Likidite Oranı} = \frac{\text{Likit Aktifler}}{\text{Kısa Vadeli Borçlar}}$$

veya

$$\text{Likidite Oranı} = \frac{\text{Hazır Değerler + Menkul Kıymetler+KVTicari Alacaklar + Diğer KV Alacaklar}}{\text{Kısa Vadeli Borçlar}}$$

Likidite oranı, piyasalarda doğabilecek bir krizde şirketlerin satışlarında olabilecek bir sıkıntı durumunda stoklar gibi kolayca nakite çevrilemeyecek kalemler göz önüne alınarak, dönen varlıkların geri kalan likit değerleriyle borç ödeyebilme kapasitesini ölçmektedir. Likidite oranı, olası bir kriz esnasında cari orana nazaran daha iyi bir gösterge olarak kabul edilmektedir. Cari oran gibi net işletme sermayesindeki değişiklikler ile birlikte değerlendirilmesi, oranın daha doğru yorumlanmasını sağlayacaktır.

Batılı finans kurumlarında, bu oranın 1 olması genel kabul gördüğü halde, yüksek enflasyonlu ve kıt fon kaynaklarına sahip ülkelerde endüstri ve sektörlerin değişik özellikleri göz önünde bulundurulduğunda asit test oranı için yeterli görülen miktar için değişiklikler söz konusudur. İşletmeler genelde kısa vadeli kaynakları kullanmak zorunda kaldıklarından bu oran genelde 1'in altında çıkmaktadır. Bu oran değerlendirilirken firmanın kısa vadeli borç yapısı da önem kazanır. Müşteri avansları ile fonlama yapan ya da peşin tahsilat yapan işletmeler ile stokları kolayca nakde çevrilebilir nitelikte olan şirketlerde oranın düşük çıkması normal karşılanabilir. Oranın 1'den büyük olması durumunda bile alacak tahsil kalitesi düşükse bu olumsuz bir durumdur. Stok devir hızı yüksek bir firmada ise oranın 1'den az olması sorun yaratmaz. Bu nedenlerle bu oranla birlikte stok devir hızı ve alacak tahsilat süresi de değerlendirilmelidir. Sermaye, duran varlıklara yatırım (veya satışı), uzun vadeli borçlanma gibi kalemlerdeki değişiklikler asit test oranında değişimlere sebep olacaktır. Bu yüzden, iki dönem arasındaki asit test oranındaki değişiklikleri de incelemek gerekmektedir (Akgüç, 1994).

Nakit Oranı (NO)

Likiditeyi daha dar anlamda ölçen bir orandır. Ticari ve ticari olmayan alacaklar dahil edilmediğinden, sadece çok kısa sürede nakite çevrilebilen dönen varlıklar kalemlerinin kısa vadeli borçları karşılama gücünü ölçer.

$$\text{Nakit Oranı} = \frac{\text{Hazır Değerler + Menkul Kıymetler}}{\text{Kısa Vadeli Borçlar}}$$

Nakit oranı bize piyasalarda veya ekonomik koşullardaki herhangi bir zorluk esnasında şirketlerin en likit varlıkları ile kısa vadeli borçlarının ilk etapta ne kadarlık kısmını geri ödeyebileceğini göstermektedir. Bu oranda likitide oranından farklı olarak, paya stoklar, akreditifler, satıcılara verilen avanslar, diğer dönen varlıklar ve en önemlisi alacaklar kalemi eklenmemiştir. Bu nedenle nakit oranı daha keskin bir ölçüt olarak kabul görür. Alacaklar tahsil edilemediği ve satışların azaldığı zor durumlarda bile firmanın borç ödeme kapasitesini açığa çıkar. Batılı finans kurumlarında, bu oranın 0,2 olması yeterli olarak genel kabul görüldüğü halde, yüksek enflasyonlu ve kıt fon kaynaklarına sahip ülkelerde endüstri ve sektörlerin değişik özellikleri göz önünde bulundurulduğunda nakit oranı için yeterli görülen miktar için değişiklikler söz konusudur. Ülkemizde işletmeler kısa vadeli kaynaklarla fonlama yapmak zorunda kaldıklarından özellikle üretim tesislerinde bu oran 0,2'nin altında kalabilmektedir (Akgüç, 1994).

Net İşletme Sermayesi – Aktifler Oranı (NSAO)

Bu oran bize şirketlerin faaliyetlerine devam etmek için gereksinim duydukları net işletme sermaye tutarının aktiflerin içindeki yüzdesel dilimini verir.

$$\text{Net İşletme Sermayesi - Aktifler Oranı} = \frac{\text{Net İşletme Sermayesi}}{\text{Toplam Aktifler}} * 100$$

veya

$$\text{Net İşletme Sermayesi - Aktifler Oranı} = \frac{\text{Dönen Varlıklar - Kısa Vadeli Borçlar}}{\text{Toplam Aktifler}} * 100$$

Bu oranla yapılan analizlerde aktiflerin toplam büyüklüğündeki değişiklikler de göz önüne alınmalıdır. Net işletme sermayesinin aktifler içindeki oranının şirket ihtiyaçları için optimal düzeyde olup olmadığını ölçmek için, sektör içerisindeki diğer şirketlerin oranları ile karşılaştırılabilir veya orandaki değişim miktarı şirketin diğer verilerindeki değişimler (karlılık, net satışlar, likidite oranlarındaki değişim miktarları..) ile karşılaştırılabilir.

Likit Aktifler – Aktifler Oranı (LAAO)

Şirketlerin, likit aktiflerinin toplam aktifler içerisindeki payını gösteren bu oran ile şirketin aktiflerinin ne derece likit olduğu anlaşılabilir. Bu oran alternatif maliyetler ve işletmenin bulunduğu sektörün nakit üretme kapasitesi göz önüne alınarak değerlendirilmelidir. Bu oranın yüksek oluşu, şirketin beklenmedik durumlardaki manevra yeteneğini arttırır.

$$\text{Likit Aktifler - Aktifler Oranı} = \frac{\text{Likit Aktifler}}{\text{Toplam Aktifler}} * 100$$

2.2.2.2 Mali Yapı Oranları

Mali Yapı Oranları işletmelerin kaynak yapısının ve uzun vadeli borç ödeme gücünün ölçülmesinde kullanılır. İşletmenin yarattığı kaynakların borç özsermaye dağılımı, aktiflerin fonlamasında ne şekilde kullanıldığı; firmaların finansal yapılarının sağlamlığı, kaynak kullanımının eniyiliği ve katma değer yaratmaktaki başarıları mali yapı oranları ile ölçülebilir.

Maddi Duran Varlıklar – Özsermaye Oranı (MDVÖO)

Maddi Duran Varlıklar-Öz Sermaye Oranı bize şirketlerin maddi yatırım tutarlarının ne kadarlık kısmının öz sermaye yoluyla finanse edilebildiğini göstermektedir. Şirketlerin yatırımlarının diğer fonlama yöntemlerine göre daha uzun vadeli ve faiz maliyeti olmayan öz sermaye yoluyla finanse ediliyor olması tercih edilen bir durumdur.

$$\text{Maddi Duran Varlıklar - Öz Sermaye Oranı} = \frac{\text{Maddi Duran Varlıklar}}{\text{Öz Sermaye}} * 100$$

Bu oranın hesaplanmasındaki amaç, maddi duran varlıkların ne kadarlık kısmının öz sermaye tarafından karşılandığını görebilmektir. Bu oranın %100'den küçük olması, şirketlerin maddi duran varlıklarının (maddi yatırımlarının) tamamının öz sermaye ile finanse edildiğini ve bu varlıkların finansmanı için ek bir yabancı kaynağa gerek kalmadığını göstermektedir. Aksi takdirde, yani oranın %100'den büyük olması, maddi duran varlıkların finansmanında yabancı kaynakların kullanıldığını, sermayenin yetersiz olduğunu ve ek sermaye finansmanına gerek duyulduğunu gösterir. Ancak sermaye yoğun teknoloji kullanan şirketlerin yatırımlarını tamamladıkları ilk yıllarda, bu oranın %100'den büyük olması normal karşılanmalıdır.

Bu oran hesaplanırken, maddi duran varlıkların net değerini kullanmak gerekir. Ancak dikkat edilmesi gereken iki husus vardır: İlki, yüksek enflasyonun olduğu ülkelerde amortisman indirilmeden bu oranın hesaplanması gerekmektedir. İkinci olarak da, varlıkların büyük bir kısmının amortismanının ayrılmış olduğu bir durumda ve/veya hızlandırılmış amortismanın kullanıldığı şirketlerde, yine amortisman indirilmeden, yani bilanço kalemlerinde brüt olarak gözüktüğü şekilde kullanılması daha anlamlı sonuçlar verecektir (Akgüç, 1994).

Maddi Duran Varlıklar – Uzun Vadeli Borçlar Oranı (MDVUBO)

Maddi Duran Varlıklar-Uzun Vadeli Borçlar oranı şirketlere uzun vadeli fon tedarik etmiş olan kreditörlerin sağlamış oldukları fonların karşılığında şirketlerin ne kadarlık sabit maddi duran varlığa (yatırıma) sahip olduğunu ölçmektedir.

$$\text{Maddi Duran Varlıklar - Uzun Vadeli Borçlar Oranı} = \frac{\text{Maddi Duran Varlıklar}}{\text{Uzun Vadeli Borçlar}} * 100$$

Bu oran aynı zamanda şirketlerin maddi duran varlıklarının ne kadarlık kısmının uzun vadeli borçlar tarafından karşılandığını göstermektedir. Bu oranın %100'den küçük olması demek, şirketlerin maddi duran varlıklarının (maddi yatırımlarının) tamamının uzun vadeli borçlar ile finanse edildiğini ve bu varlıkların finansmanı için kısa vadeli yabancı kaynaklara veya ek bir sermaye tutarına ihtiyaç olmadığını göstermektedir. Oranın %100'den büyük olması durumunda, eğer aradaki fark öz sermaye kaynakları tarafından tam olarak karşılanamaz ise kısa vadeli yabancı kaynakların kullanıldığını, sermayenin ve uzun vadeli fonların yetersiz olduğunu ve uzun vadeli kaynaklarla finansmana gerek duyulduğunu gösterir. Sermaye yoğun şirketlerde oranın %100'ün üzerinde kalması doğal karşılanabilir. Oranın Maddi Duran Varlıklar-Öz Sermaye oranı ve Maddi Duran Varlıklar-Devamlı Sermaye Oranı ile birlikte değerlendirilmesi daha doğru tespitlerin yapılmasını sağlayacaktır.

Değerlendirme yapılırken ayrıca maddi duran varlıkların gerçek değeri, teknolojik gelişmeler sonucunda bu varlıkların demode hale gelmesi, firmanın likitide sorunu olup olmadığı göz önüne alınmalıdır. Bu oran hesaplanırken, maddi duran varlıkların net değerinin kullanılması gerekir. Ancak dikkat edilmesi gereken iki husus vardır: İlki yüksek enflasyonun olduğu ülkelerde amortisman indirilmeden bu oranın hesaplanma gereğidir. İkinci olarak da varlıkların büyük bir kısmının amortismanının ayrılmış olduğu bir durumda ve/veya hızlandırılmış amortismanın kullanıldığı şirketlerde, yine amortismanın indirgenmeden, yani bilanço kalemlerinde brüt olarak gözüktüğü şekilde kullanılması daha anlamlı sonuçlar verecektir (Akgüç, 1994).

Maddi Duran Varlıklar – Devamlı Sermaye Oranı (MDVDSO)

Bu oranının hesaplanması sonucu, şirketlerin maddi duran varlıklarının ne kadarlık kısmının uzun vadeli kaynaklar tarafından finanse edildiğini görebiliriz. Bu oran sayesinde şirketlerin maddi yatırımlarını fonlamada ne kadar başarılı olduğunu ölçebiliriz.

$$\text{Maddi Duran Varlıklar - Devamlı Sermaye Oranı} = \frac{\text{Maddi Duran Varlıklar}}{\text{Devamlı Sermaye}} * 100$$

veya

$$\text{Maddi Duran Varlıklar - Devamlı Sermaye Oranı} = \frac{\text{Maddi Duran Varlıklar}}{\text{Uzun Vadeli Borçlar + Öz Sermaye}} * 100$$

Maddi duran varlıkların tamamı öz sermaye tarafından karşılanmayan şirketlerde bu oranın hesaplanması önem kazanır. Böylece, uzun vadeli varlıkların, en azından hangi oranda uzun vadeli kaynaklarla karşılanıyor olduğunu görürüz.

Bu oranın %100'den küçük olması, şirketlerin maddi duran varlıklarının (maddi yatırımlarının) tamamının uzun vadeli kaynaklarla ve öz sermayeleri ile finanse edildiğini ve bu varlıkların finansmanı için likitide ve faiz riski teşkil eden kısa vadeli borçlara gerek kalmadığını göstermektedir. Oranın %100'den büyük olması durumunda, maddi duran varlıkların finansmanında kısa vadeli yabancı kaynakların kullanıldığını, devamlı sermayenin yetersiz olduğunu ve ek bir fon ihtiyacına gerek duyulduğunu anlayabiliriz. Değerlendirme yapılırken maddi duran varlıkların gerçek değeri, teknolojik gelişmeler sonucunda bu varlıkların demode hale gelebilmesi, firmanın likitide sorunu olup olmadığı göz önünde alınmalıdır.

Ancak yoğun teknoloji kullanan şirketlerin yatırımlarını tamamladıkları ilk yıllarda, bu oranın %100'den büyük olması normal karşılanmalıdır.

Bu oran hesaplanırken, maddi duran varlıkların net değerinin kullanılması gerekir. Ancak dikkat edilmesi gereken iki husus vardır: İlki yüksek enflasyonun olduğu ülkelerde amortisman indirilmeden bu oranın hesaplanma gereğidir. İkinci olarak da, varlıkların büyük bir kısmının amortismanının ayrılmış olduğu bir durumda ve/veya hızlandırılmış amortismanın kullanıldığı şirketlerde, yine amortismanın indirgenmeden, yani bilanço kalemlerinde brüt olarak gözüktüğü şekilde kullanılması daha anlamlı sonuçlar verecektir (Akgüç, 1994).

Duran Varlıklar – Uzun Vadeli Borçlar Oranı (DVUBO)

Duran Varlıklar-Uzun Vadeli Borçlar oranı şirketlere uzun vadeli fon tedarik etmiş olan kreditörlerin sağlamış oldukları fonların karşılığında şirketlerin ne kadarlık sabit duran varlıklara (yatırımlara) sahip olduğunu göstermektedir. Bu oran şirkete kredi verenler tarafından bir emniyet payı olarak kullanılmaktadır.

$$\text{Duran Varlıklar - Uzun Vadeli Borçlar Oranı} = \frac{\text{Duran Varlıklar}}{\text{Uzun Vadeli Borçlar}} * 100$$

Bu oranın %100'den küçük olması, şirketlerin duran varlıklarının (sabit yatırımlarının) tamamının uzun vadeli borçlar ile finanse edilebildiğini ve bu varlıkların finansmanı için kısa vadeli yabancı kaynaklara veya ek bir sermaye tutarına ihtiyaç olmadığını göstermektedir. Aksi takdirde, yani oranın %100'den büyük olması durumunda, eğer aradaki fark öz sermaye kaynakları tarafından tam olarak karşılanamaz ise kısa vadeli yabancı kaynakların kullanıldığını, sermayenin ve uzun vadeli fonların yetersiz olduğunu anlayabiliriz. Bu daha fazla uzun vadeli kaynağa ihtiyaç duyulduğunun belirtisidir. Oranın Duran Varlıklar-Öz Sermaye oranı ve Duran Varlıklar-Devamlı Sermaye oranı ile birlikte değerlendirilmesi daha doğru tesbitlerin yapılmasını sağlayacaktır.

2.2.2.3 Kaldıraç Oranları

Kaldıraç oranları, işletmenin kaynak yapısını gösterir ve işletmenin varlıklarının hangi kaynaklarla ne oranlarda karşılandığını görebilmemizi sağlar. Bu oranlar ile işletmenin uzun vadeli borçları ödeme gücü gibi mali yapı durumu hakkında fikir sahibi olabiliriz.

Sermaye – Özsermaye Oranı (SÖSO)

Bu oranın yüksek olması esas itibariyle öz sermayenin ödenmiş sermaye yoluyla sağlandığı, yani, ortakların sermaye eklemesi ile oluştuğu, karların büyük miktarda dağıtıldığı veya sermayeye eklendiği (bedelsiz sermaye artırımı), esasen şirketin yüksek karlar elde etmediği sonucu ortaya çıkar.

$$\text{Sermaye - Öz Sermaye Oranı} = \frac{\text{Sermaye}}{\text{Öz Sermaye}} * 100$$

Diğer yandan bu oranın düşük çıkması, şirketin oto-finansmana önem verdiği, yani karlarını fazla dağıtmayarak bünyesinde (yedekler hesabında) tuttuğu sonucunu ortaya çıkarır.

Burada dikkat etmemiz gereken husus ise, bedelsiz sermaye artırımı yoluyla öz sermayeyi oluşturan diğer alt kalemlerin ödenmiş sermaye kalemine transfer edilmiş olabileceğidir. Oran bu sebepten dolayı yüksek çıkabilir. Bu sebepten dolayı oranı incelerken son zamanlarda bedelsiz sermaye artırımı yapılıp yapılmadığına dikkat edilmelidir.

Uzun Vadeli Borçlar – Devamlı Sermaye Oranı (UBDSO)

Uzun Vadeli Borçlar ile Öz Sermayenin yani şirketlerin uzun vadeli kaynak toplamını gösteren devamlı sermayenin içerisindeki uzun vadeli borçların ağırlığını göstermekte olan bu oran, şirketlerin uzun vadeli sermaye yapısı ve politikaları hakkında bilgi vermektedir.

$$\text{Uzun Vadeli Borçlar- Devamlı Sermaye Oranı} = \frac{\text{Uzun Vadeli Borçlar}}{\text{Devamlı Sermaye}} * 100$$

veya

$$\text{Uzun Vadeli Borçlar- Devamlı Sermaye Oranı} = \frac{\text{Uzun Vadeli Borçlar}}{\text{Uzun Vadeli Borçlar} + \text{Öz Sermaye}} * 100$$

Bu oranın, genel bir kural olarak, sanayi işletmelerinde %33'ü, kamu hizmeti veren işletmeler de ise %50'yi aşmaması gerektiği söylenebilir. Ticari işletmeler genellikle uzun vadeli kaynaklardan daha az yararlandıkları için bu oran özellikle sanayi şirketleri için kullanılmaktadır (Akgüç, 1994).

Uzun Vadeli Borçlar – Aktifler Oranı (UBAO)

Toplam varlıkların ne kadarlık kısmının uzun vadeli yabancı kaynaklardan fonlandığını göstermekte olan bu oran, şirketlerin uzun vadeli fon temin edebilme gücü hakkında da bilgi verebilmektedir.

$$\text{Uzun Vadeli Borçlar- Aktifler Oranı} = \frac{\text{Uzun Vadeli Borçlar}}{\text{Toplam Aktifler}} * 100$$

Oranın yüksek olması, şirketlerin varlıklarını kolayca uzun vadeli kaynaklardan fonlayabildiğini göstermekle beraber, oranın aşırı yüksek olması özellikle durgunluk dönemlerinde firmanın borç taksitlerini ödemede zorlanabileceğinin göstergesidir. Genelde büyük yatırımlara başlamış olan şirketlerde bu oran yüksektir. Yatırımlardan sonra uzun dönemde bu oranın düşmemesi yatırımların yeteri kadar katma değer getiremediği şeklinde yorumlanabilir. Sektörel analizlerde de dikkat edilmesi gereken bir orandır. Ayrıca oranın daha iyi yorumlanabilmesi için, kısa vadeli borçların da izlenmesi daha doğru yargılara varılmasını sağlayacaktır.

Kısa Vadeli Borçlar – Toplam Borçlar Oranı (KBTBO)

Yabancı kaynakların vadelerine göre dağılımı hakkında fikir vermekte olan borç yapısı oranı, yabancı kaynaklar içerisindeki kısa vadeli borçların ağırlığı hakkında bilgi verir.

$$\text{Kısa Vadeli Borçlar- Toplam Borçlar Oranı} = \frac{\text{Kısa Vadeli Borçlar}}{\text{Toplam Aktifler}} * 100$$

Genelde %75 civarında olması beklenen bu oran zaman zaman uzun vadeli kaynak bulmanın zorlaştığı ekonomik şartlarda %90 - %100'lere kadar çıkmaktadır. Dönen Varlıklar-

Aktifler oranı yüksek ya da emek yoğun olan firmalarda yapıları gereği bu oran nispeten daha yüksektir. Duran varlık oranı daha yüksek sermaye yoğun firmalarda bu oranın göreceli olarak daha aşağı olması beklenir. Belli bir zaman diliminde orandaki gelişimler bize şirketin risk durumu hakkında da bir gösterge olabilmektedir. Oranın yükselmesi şirketlerin uzun vadeli kaynak bulmada zorluklarının bulunduğunu göstermektedir ve şirketlerin risklerinin artması anlamına gelmektedir (Akgüç, 1994).

Özsermaye Çarpanı (ÖÇ)

Finansal kaldıraç ile benzer neticeler veren öz sermaye çarpanı, aktiflerin öz sermayenin kaç katı olduğunu, başka bir deyişle öz sermayenin aktifler içerisindeki ağırlığını göstermektedir.

$$\text{Borç- Öz Sermaye Oranı} = \frac{\text{Toplam Aktifler}}{\text{Öz Sermaye}}$$

Öz sermaye çarpanının %200’den küçük olması gereği batılı finans kuruluşlarında genel bir kural olarak kabul edilmiş olmasına rağmen sermaye arzının düşük olduğu ülkemizde ilgili oran %250 civarındadır. Öz sermaye çarpanının düşük olması yani işletmenin iktisadi varlıklarının göreceli olarak daha fazla öz sermaye ile fonlanması kreditörler için düşük risk anlamına gelir. Fakat satışlarda kar marjı yüksek olan şirketler optimal yabancı kaynak kullanımıyla öz sermaye karlılığını artırabilirler. Oranın bu şekilde artması şirketin riskinin artması şeklinde yorumlanabileceği gibi, ortakların sermaye başına daha fazla kar payı (temettü) elde ettiği (kaldıraç etkisi) anlamına da gelmektedir (Akgüç, 1994).

Borç-Özsermaye Oranı (BÖO)

Firmanın aktiflerini fonlamada özkaynaklarına oranla ne kadar yabancı kaynak kullanıldığını gösteren Borç-Öz Sermaye oranının düşük çıkması tercih edilmektedir.

$$\text{Borç- Öz Sermaye Oranı} = \frac{\text{Uzun Vadeli Borçlar+ Kısa Vadeli Borçlar}}{\text{Öz Sermaye}} * 100$$

Bununla birlikte, oranı oluşturan etmenlerin iyi bir şekilde incelenmesi gerekmektedir. Şirketlerin sermaye yapıları hakkında iyi bir gösterge olan Borç-Öz Sermaye oranının, zaman içerisinde göstereceği gelişmeler şirketlerin sermaye yapılarındaki tercihleri hakkında gözlem yapma olanağını vermektedir. Borç-Aktifler oranında da bahsedildiği gibi özkaynaklar ile yabancı kaynaklar arasında uygun bir dengenin kurulması oldukça önemlidir. Uygun yönetim anlayışı ile çalışan ve içinde bulunduğu sektörün elverdiği firmalar iş riski, faaliyet riski oranları düşük ise daha yüksek Borç-Aktifler ve Borç-Öz Sermaye oranları ile çalışabilirler. Oranın %100’ü aşmaması gerektiği batılı finans kurumları tarafından genel bir kural olarak benimsenmişken, ülkemiz gibi öz sermaye bulmakta zorluk çekilen ülkelerde bu oranın %150 - %200 arası olması kabul edilebilir (Akgüç, 1994).

Borç-Aktifler Oranı (BAO)

Toplam aktiflerin ne kadarlık kısmının yabancı kaynaklardan fonlandığını göstermekte olan Borç-Aktifler oranı aynı zamanda şirketlerin toplam kaynaklarının dağılımı hakkında da bilgi vermekte olup, borçlanma katsayısı ile benzer sonuçlar vermektedir.

$$\text{Borç-Aktifler Oranı} = \frac{\text{Uzun Vadeli Borçlar+ Kısa Vadeli Borçlar}}{\text{Toplam Aktifler}} * 100$$

Şirketlerin hem toplam varlıklarının ne kadarlık kısmının yabancı kaynaklardan fonlandığı, hem de kaynakların niteliklerine göre dağılımı hakkında fikir veren Borç-Aktifler oranı, şirket için bir risk göstergesi olarak da kullanılabilir. Bu oranın yüksek çıkması işletmeyi kredi verenler açısından riskli pozisyona düşürmektedir. Fakat yabancı kaynak kullanımı belli bir optimaliteye kadar öz sermaye karlılığını artıran bir durumdur. Bu optimalite noktası ortalama kaynak maliyetinin minimum olduğu noktadır. Oran yüksek olsa da borçlanma maliyeti üzerinde geliştirilebilen katma değer, yönetimin az bir öz sermaye ile geniş bir kaynağı yönlendirebildiğini gösterir. Ortaklara daha düşük sermaye ile sermaye başına daha fazla kar payı (temettü) alma olanağı (kaldıraç etkisi) sağlamaktadır. Buna rağmen kaldıracın arttırılmasının riski arttırdığı göz ardı edilmemelidir. Finansal kaldıraç oranının %50'den küçük olması gereği batılı finans kuruluşlarınca genel bir ölçüt olarak benimsenmiş iken, ülkemiz gibi sermaye arzının düşük olduğu ülkelerde ilgili oranın %60'tan yüksek olması normal karşılanmaktadır. Bu oran değerlendirilirken firmanın yeniden değerleme yapıp yapmadığına da bakılmalıdır. Yeniden değerleme sonucu oran düşük izlenimi verebilir (Akgüç, 1994).

2.2.2.4 Piyasa Değerini Ölçen Oranlar

Halka açık şirketlerde ortakların yatırımları sonucu elde ettikleri getirilerini görmek ve gelecekteki olası getirilerini tahmin etmek, hisse senetlerinin piyasada oluşan fiyatlarının gerçekçiliğini ölçmek için kullandıkları oranlardır.

Piyasa Değeri (PD)

Şirketlerin hisse senetlerinin İMKB'de oluşan fiyatı ve toplam hisse senedi sayısı ile hesaplanan bu değer şirketlerin piyasa tarafından belirlenen değerini göstermektedir. Piyasa değeri, şirketlerin borsadaki senetlerin alım-satım işlemleri sırasında almış olduğu değere göre değişmektedir.

Piyasa değerini belirleyen etmenlerden biri hisse senedi sayısıdır. Sermaye artırımı yoluyla hisse senetlerin artması durumunda piyasa değeri sadece bedelli sermaye artırımı esnasında kullanılan toplam rüçhan hakları miktarı kadar artmaktadır(Akgüç, 1994). Formülü,

$$\text{Piyasa Değeri} = \frac{\text{Hisse Senedi Fiyatı} * \text{Ödenmiş Sermaye}}{1000}$$

şeklinde verilebilir.

Hisse Başına Kar (HBK)

Net dönem karının ödenmiş sermayeye bölünmesi ile bulunan hisse senedi başına kar, bir şirketin piyasa değerini belirleyen en önemli etmenlerden biridir. Bir yatırımcı bu oran sayesinde sahip olduğu her bir hisse senedi başına şirketin karından ne kadarlık pay düştüğünü görebilmektedir.

Şirketlerin faaliyet nedenini oluşturan ve ortaklara bir getiri sağlayabilmek amacıyla sağlanması gereken net dönem karının beher hisse senedi başına olan payını ölçen bu oran, hisse senetlerin fiyatını ölçmede kullanılan Fiyat-Kazanç oranı içinde ayrıca bir veri olarak kullanılmaktadır(Akgüç, 1994). Formülü,

$$\text{Hisse Başına Kar} = \frac{\text{Net Dönem Karı}}{\text{Ödenmiş Sermaye}} * \text{Kar}$$

şeklinde verilebilir.

Fiyat-Kazanç Oranı (F/K)

F/K oranı sayesinde şirketlerin beher hisse senedi başına elde edeceği kara karşılık, hisse senedini almak için ne kadar ödememiz gerektiğini bulabiliriz. F/K oranının yüksekse hisse fiyatının yüksek olduğu, oranın düşükse hisse fiyatının düşük olduğu şeklinde yargılara varılabileceği gibi benzer şirketler ve endüstri/sektör ortalamasına bakarak yorumlar yapılabilir.

F/K oranı bize ayrıca hisse senedi başına kar rakamı sabit kalacağı ve elde edilen karların tamamının dağıtılacağı varsayılarak, hisse senedi için ödeyeceğimiz fiyatın bize temettü olarak kaç yılda geri döneceğini gösterir. Bu aynı zamanda şirkete olan güvenin bir göstergesi olarak da kullanılabilinir. Yüksek bir F/K oranı şirketin piyasa fiyatının şişmiş olduğunu gösterir. Ancak bukanıya varabilmek için şirketin F/K oranının yanında, bulunduğu sektörün ortalamaF/K oranı ve piyasa genelinin F/K oranının da çok iyi incelenmesi gerekmektedir(Akgüç, 1994). Formülü,

$$\text{F/K Oranı} = \frac{\text{Hisse Senedi Fiyatı}}{HBK_{Şirket}} = \frac{\text{Hisse Senedi Fiyatı} * \text{Ödenmiş Sermaye}}{\text{Net Dönem Karı} * 1000}$$

şeklinde verilebilir.

Piyasa Değeri-Defter Değeri Oranı (PD/DD),

PD/DD, şirketlerin piyasada oluşan değerlerinin, özsermayelerinin kaç katı olduğunu gösterir. Defter değeri, bir şirketin hisse senetlerinin piyasa değerinin arkasında, hisse senetleri için güvence oluşturan maddi varlığının değerini gösterir.

PD/DD oranı sayesinde şirketlerin beher hisse senedi başına düşen özsermaye tutarına karşılık, hisse senedini almak için ne kadar ödememiz gerektiğini göstermektedir. PD/DD oranı yüksek olduğunda, hisse fiyatının yüksek olduğu, düşük olduğunda ise hisse fiyatının düşük olduğu şeklinde yargılara varılabileceği gibi benzer şirketler ve endüstri / sektör ortalamasına bakarak yorumlar yapılabilir(Akgüç, 1994). Formülü,

$$\text{PD/DD Oranı} = \frac{\text{Piyasa Değeri}}{\text{Öz Sermaye}}$$

$$\text{Hisse Başına PD/DD Oranı} = \frac{\text{Hisse Fiyatı}}{\text{Hisse Başına DD}}$$

şeklinde verilebilir.

Fiyat-Nakit Akışı Oranı (F/NK)

F/NA, şirketlerin piyasada oluşan değerlerinin, şirkette devrelik olarak gerçekleşen nakit akışlarının kaç katı olduğunu gösterir. Formülü,

$$\text{Fiyat/Nakit Akışı Oranı = F/NA Oranı} = \frac{\text{Piyasa Değeri}}{\text{(Net Kar-Zarar+Amortisman)}}$$

şeklinde verilebilir.

Bir işletmenin finansal riskini değişik yaklaşımlarla ölçmek mümkündür. Bu yaklaşımlarda yüzlerce finansal oran göz önüne alınabilir. Portföy yaklaşımı, varyans analizi ve klasik finansal oran analizi bu amaçla kullanılan yaklaşımlardan bazılarıdır (Aktaş v.d., 2003). Bunlar içerisinde uygulamada en çok kullanılanı klasik finansal oran analizi olup, yapısı itibariyle tek boyutlu bir analizdir. Bu analizde, bir işletmenin mali performansını ortaya koyabilmek için finansal oranlar teker teker ele alınır ve bunun neticesinde işletmenin mali performansı bir başka deyişle finansal riski hakkında genel bir kanaate ulaşılmaya çalışılır. Klasik mali oran analizi bu yapısından dolayı birtakım eksiklikler taşır (Aktaş v.d., 2003). Söz konusu eksiklikleri ortadan kaldırmak için mali oranların bir bütün olarak ele alındığı çok boyutlu finansal oran analizi günümüzde yaygın olarak kullanılmaktadır (Akgüç, 1994, Aktaş v.d, 2003, Ercan ve Ban, 2005, Yalçıner v.d., 2005). Çok boyutlu finansal oran analizi, "ölçüt seçimi", "seçilen ölçütlere verilen ağırlık" ve "sonuçların yorumlanması" öznel veya nesnel olmaya bağlı olarak iki başlık altında toplanmaktadır. Bu üç başlıkta nesnel olan yaklaşıma günlük kullanımda "riskmetre" denilirken, öznel yaklaşımlar daha çok "istatistiksel yaklaşım", "matematiksel sınıflandırma yaklaşımı" ve "yapay sinir ağı yaklaşımı" başlıkları altında yer almaktadır (Aktaş v.d., 2003). Literatürde çok boyutlu modellerde en sık kullanılan istatistiksel yöntemler; çoklu regrasyon modeli, diskriminant analizi ve logit modelleridir (Aktaş v.d., 2003, Johnson and Wichern, 1992).

2.3 Finansal Performansı Değerlendirilecek Şirketlerin ve Zaman Aralığının Belirlenmesi ve Finansal Oranların Hesaplanması

On dokuz adet finansal oranı hesaplanacak şirketlerin belirlenmesinde, bilânçoların benzerliği göz önüne alınmıştır. Bazı şirketler için bilânçolarda ortak kalemlerin olmadığı gözlenmiştir. Örneğin hizmet sektöründe yer alan şirketler için stok kaleminin sıfır olması nedeniyle bu şirketler için bazı finansal oranların hesaplanması mümkün olmamaktadır. Finansal oranların hesaplanmasında İMKB'deki şirketlerin bir alt kümesi olan Ulusal Sınaî Endeksi'nde yer alan şirketler kullanılmıştır. Ulusal Sınaî Endeksi'nde yer alan şirketler "XUSIN" adı altında İMKB'nin resmi internet sitesinde verilmektedir (http://www.imkb.gov.tr/endeksler.htm). Finansal oranların hesaplanmasında zaman aralığı, 2002–2007 yılları arası olarak belirlenmiştir. İMKB tarafından 2007 yılında yayınlanan listeye göre belirlenmiş olan Ulusal Sınaî Endeksi'ndeki şirketlerin 2002–2007 yılları arasındaki bilânço dosyaları, tüm dosyalar içinden süzülmüştür. 2007 yılında bu endekste 151 şirket olmasına rağmen, 2002–2007 yılları arasında verileri tam olan şirket sayısı 132 adet olarak belirlenmiş ve çalışmada kullanılmak üzere bu şirketler seçilmiştir. Bu 132 şirketin her biri için 12 aylık bilânçolarında verilen bilgiler yardımıyla 19 finansal oran hesaplanmıştır. Böylece şirketler için 132×133 boyutunda ($133 = 19x7$) bir özellik matrisi elde edilmiştir.

Şirket bilânçolarında standart bir format olmayışı nedeniyle finansal oranların hesaplanması faaliyeti, basit bir matematiksel işlem olmasına rağmen, en çok çalışma zamanı tüketen faaliyet olmuştur. Özellikle 2003 ve 2004 yıllarına ait şirket bilânçolarından birçoğunda finansal oranların hesaplanması, elle yapılmıştır. Finansal oranları hesaplamayı tamamen otomatik hale getirmek mümkün gözükmemektedir. Ancak çalışmalarımız kapsamında finansal oranların, bir bilgisayar programı aracılığı ile hesaplanabilmesi için önemli aşamalar kat edilmiştir. Belirlenen en yaygın bilânço şablonları için Excel-VBA programı geliştirilmiştir. Bilânçolardaki kalemlerde oluşan isim farklılıkları bir uzman desteği ile giderilmeye çalışılmıştır. Örneğin 2005–2007 arasında bilânçolarda yer alan "Toplam Varlıklar" kalemi 2002–2004 yılları arasında "Toplam Aktifler" ismiyle yer almaktadır. Bu farklılıkların yanı sıra büyük-küçük harf farklılıkları da göz önüne alındığında ortak bir program yazmak ve ortak bir format oluşturmak oldukça zorlaşmaktadır. Ancak benzerlikler ve farklılıklardan yararlanılarak yazılan programlar sayesinde ilgilenilen dönemler için verilerin büyük bir bölümü program yardımıyla hesaplanabilmiştir. Finansal oranları hesaplanan toplam 924 bilânçonun 834'ü bu programlar yardımıyla kalan 90'ı ise el ile gerçekleştirilmiştir.

Geliştirilen bilgisayar programının temel adımları aşağıda verilmiştir.

Adım 1. Excel-VBA Programı, bilgisayardaki yeri tanımlanmış Excel sayfasında ilk sütunda liste olarak verilen şirket isimlerine göre sıradaki şirkete ait bilânço dosyasını açar,

Adım 2. Önceden belirlenen finansal oranları hesaplamak için gerekli olan tüm değerleri uygun formattaki bilânçodan okur ve finansal oranları hesaplar,

Adım 3. Hesaplanan oranları ve şirket ismini diğer bir dosyaya yazar ve bilânço dosyasını kapatır,

Adım 4. Sırada başka bir şirket varsa onun için Adım 1'e dönülür, aksi halde program sonlanır.

Adım 5. Program sonlandığında hesaplanamayan oranlar tekrar gözden geçirilerek tablo tamamlanır.

Yazılan program ve elle yapılan hesaplamalar sonucu Ulusal Sınaî Endeksi'nde yer alan 132 şirket için oluşturulan finansal oranlar, 2002 yılından başlamak üzere EK Tablo 1 – Ek Tablo 6 ile verilmiştir.

Bu finansal oranlar kullanarak 132 adet şirketin, hisse senedi getirilerine ve karlılıklarına göre sınıflandırılması yapılmıştır.

Yukarıda da bahsedildiği gibi çok sınıflı sınıflandırma problemlerinin çözümü genellikle iki sınıflı sınıflandırma problemlerinin çözüm yöntemlerine dayanmaktadır. Bu projede ikili sınıflandırma problemleri için geliştirilen polihedral konik fonksiyonlar (PKF) temelli iki farklı yaklaşımın kullanılması hedeflenmektedir. Bunlardan birincisi PKF algoritması, ikincisi ise detayları aşağıda açıklanacak olan iki amaçlı tamsayılı matematiksel modele dayalı sınıflandırma yaklaşımıdır.

Proje kapsamında yürütülen çalışmalar sonucunda Polihedral Konik Fonksiyonlar kullanılarak tamsayılı programlama yaklaşımı ile yeni bir ikili sınıflandırma yaklaşımı geliştirilmiştir.

3 YENİ ÖNERİLEN İKİ VE ÇOK SINIFLI SINIFLANDIRMA YÖNTEMLERİ

3.1 Polihedral Konik Fonksiyonlar Algoritması: Analiz ve Motivasyon

A ve B kümeleri Rn'de verilmiş iki küme olmak üzere

$$A = \left\{a^i \in R^n : i \in I\right\}, \qquad I = \{1,\ldots,m\}$$
$$B = \left\{b^j \in R^n : j \in J\right\}, \qquad J = \{1,\ldots,p\}$$

olsun. Bölüm 2.1.1.4'de açıklanış olan PKF Algoritmasının her ardıştırmasında A kümesinin keyfi bir *a* noktası seçilir ve bir doğrusal programlama problemi çözülerek bulunan

$$w \in R^n,\ \xi \in R^+ = [0,+\infty)\ ve\ \gamma \ge 1$$

parametrelerine bağlı, grafiği bir koni, ve seviye kümesi de bir dışbükey polihedron olan

$$g_{(w,\xi,\gamma)} : R^n \to R$$

polihedral konik fonksiyonu

$$g_{(w,\xi,\gamma)} = w'(x-a) + \xi\|x-a\|_1 - \gamma$$

şeklinde tanımlanır.

Bu fonksiyonun epigrafı, tepe noktası

$$(a, -\gamma)$$

noktasında olan dışbükey bir konidir. Oluşturulmuş olan doğrusal programlama problemi, bu fonksiyonun parametrelerinin, *B* kümesinin hiçbir noktasını, *A* kümesinden ise alabildiğince çok sayıda noktayı epigrafın içine alabilecek şekilde tanımlanmasını sağlamaktadır. Hatırlatalım ki, *g* fonksiyonunun epigrafı

$$epi(g) = \{(x,\alpha) \in R^n \times R : g(x) \le \alpha\}$$

şeklinde tanımlanmaktadır. Böylece, algoritmanın her ardıştırmasında, *A* kümesinin *B*'den ayrılan noktaları *A* kümesinden çıkartılarak geriye kalan kısmı için yeni bir ayırıcı polihedral konik fonksiyon oluşturulur ve bu süreç, *A* kümesinde *B*'den ayrılmamış nokta kalmayana kadar devam eder. Nihai ayırıcı fonksiyon, her ardıştırmada oluşturulmuş olan bütün polihedral konik fonksiyonların noktasal minimumu şeklinde tanımlanır.

Bu şekilde tasarlanmış olan PKF Algoritması herhangi bir veri kümesinde 100% eğitim başarısı sağlamaktadır. Bu, çok önemli bir gösterge olsa da ve her ne kadar birçok veri kümeleri için yüksek başarı sonuçları elde edilse de bazı durumlarda test kümelerinde istediğimiz başarı oranı yakalanamamaktadır. Dahası, bazı durumlarda elde edilen eğitim kümesindeki başarı oranı ile test kümesindeki başarı oranları arasındaki farklar, bu algoritmanın genelleme özelliğinin her zaman istenilen seviyede olmadığını göstermektedir.

PKF algoritmasının adımlarından açık bir şekilde göründüğü gibi ve çok sayıda test problemleri üzerinde yapılan denemelerin de gösterdiği gibi, her ardıştırmada tanımlanan ayırıcı polihedral fonksiyonların "ayırma performansı" ve bunlara bağlı olarak tanımlanan nihai ayırıcı fonksiyonun başarı oranı büyük ölçüde tepe noktasının "keyfi" seçilen birleşeni olan "*a*" noktasına çok bağlıdır. Yaptığımız çok sayıda denemeler sonucunda, diğer parametrelerin yanı sıra, bu noktalar da uygun bir matematiksel modelin optimal çözümü olarak bulunduğunda ortaya çıkan polihedral konik fonksiyonların, veri madenciliği açısından daha yüksek performansa sahip ayırıcı fonksiyonlar oluşturma olanağı sağladığı görülmüştür.

Aşağıda detaylarını açıklayacağımız çok amaçlı (iki amaçlı) tam sayılı matematiksel modele dayanan yeni bir algoritmanın ana fikri buradan alınmıştır. Önce, belli sayıda "tepe noktası birleşeni" seçilerek her nokta için yukarıdaki algoritmada tanımlanan problem bir kez çözülüyor. Böylece, belli sayıda polihedral konik fonksiyon elde edilir. Bu fonksiyonların her biri, *A* kümesinden belirli sayıda noktayı *B* kümesinden ayırmaktadır. Nihai ayırıcı fonksiyon, bu fonksiyonların noktasal minimumu şeklinde değil, daha farklı bir şekilde tanımlanıyor. İki amaçlı tam sayılı problem, en az sayıda fonksiyon kullanılarak olabildiğince çok sayıda noktayı *B* kümesinden ayıracak şekilde tanımlanmaktadır. Ayrıca, fonksiyonu tanımlayacak problem çözülürken, PKF Algoritmasında olduğu gibi "her adımda B kümesinden hiç bir noktaya izin verilmemesi" kuralı da bozulmuş ve B kümesi elemanlarına da oluşturulacak polihedrona girme şansı tanıyan belli bir "hata" oranına izin verilmiştir. Bunun sayesinde daha esnek ve daha az "açgözlülü" bir yöntem elde edilmiştir. Böylece, yeni sınıflandırma yöntemi daha fazla genelleme özelliğine sahip olmaktadır. Oluşturulacak olan parametre matrisine uçu açık bir şekilde genişleyebilme özelliğinin tanınması da yöntemin başarı oranının yükseltilmesi için uçu açık bir şans tanınması gibi değerlendirilebilir.

3.2 İki amaçlı tam sayılı matematiksel programlama temelli algoritma

Sadece tek bir polihedral konik fonksiyon ile A kümesinin elemanlarının tümünü B kümesinden ayırmak uç bir durumdur. Böyle bir fonksiyon olsa bile, o fonksiyonun epigrafının tepe noktasının nerede olduğunun belirlenmesi de önemli bir problemdir. Doğrusal programlama problemlerinin hızlı bir şekilde çözülebilmesi sebebiyle tüm

$$a_i \in A$$

noktalarını merkez olarak belirleyip m adet ayırıcı polihedral konik fonksiyonun elde edilmesi zor değildir. Böyle bir süreç ile A kümesinin hangi noktalarını ayırdığı bilinen m tane (veya daha az veya daha fazla sayıda) polihedral konik fonksiyon elde edilebilir. Amacımız bu fonksiyonların içinden belirli (optimal) sayıda fonksiyonu seçerek iki kümenin en başarılı şekilde ayrılmasını sağlamaktır. Bunun için iki amaçlı tamsayılı programlama modeli önerilmektedir. Önerilen iki amaçlı yapı, ağırlıklı toplam ve konik skalerleştirme yöntemleri kullanılmakla tek amaçlı hale dönüştürülerek literatürde iyi bilinen Liver, WBCP, Heart, Ionosphere ve Diabet veri kümeleri üzerinde denenmiştir.

Başlangıç adım. a^l merkezli, *A* kümesinden olabildiğince çok sayıda noktayı içine alan ve *B* kümesinden hiçbir noktayı içine almayan (ileride B kumesinden de bazı elemanları alabilecek şekilde bir tolerans parametresi belirlenecektir) bir adet PKF'nin oluşturulması için aşağıda tanımlanan P_l problemi çözülür. Bu problemde tanımlanan birinci grup kısıtlar *A* kümesinde yer alan tüm noktalar için yazılır. PKF fonksiyonunun oluşturduğu koninin seviye kümesi ile elde edilen dışbükey kümenin içi yani fonksiyonun seviye kümesinin negatif kısmında kalan *A* kümesine ait elemanlar için *y* değişkeni 0 değeri alır. Bu durumda, *y* değişkeninin sıfır değerine tekabül eden noktalar, oluşturulmuş olan bu fonksiyon ile *B* kümesinden ayrılmış olur. İkinci grup kısıtlar ise *B* kümesinde yer alan noktalar için yazılır ve bu kısıtlar tüm *B* kümesine ait noktaların, oluşturulacak olan fonksiyonun seviye kümesinin pozitif kısmında yer almasını sağlar. Burada amaç, birinci grup kısıtta mümkün olduğunca çok sayıda *A* kümesine ait noktanın koninin içinde kalmasını sağlayacak şekilde *A* kümesindeki her noktaya karşı gelen *y* değişkenlerinin ortalamasının en küçüklenmesidir.

Böylece, açıkladığımız model aşağıdaki şekilde yazılabilir:

$$w'(a^i - a^l) + \xi\left\|a^i - a^l\right\|_1 - \gamma + 1 \le y, \quad \forall i \in I$$

$$-w'(b^j - a^l) - \xi\left\|b^j - a^l\right\|_1 + \gamma + 1 \le 0, \quad \forall j \in J$$

$$w \in R^n, \ \ \xi \in R^+ = [0, +\infty), \ \ \gamma \ge 1, e_m = [1,1,....,1]^T_{1xm}$$

Kısıtları altında

$$\min\left(\frac{y' e_m}{m}\right) \qquad \textbf{(}\boldsymbol{P_l}\textbf{)}$$

3.2.1 İki amaçlı tam sayılı matematiksel model

A kümesindeki her $a_1, a_2, \ldots, a_m$ noktası merkez olarak kabul edilerek, (a^l=a_i), P_l probleminin çözümü ile elde edilen PKF'ler sırasıyla $g_1, g_2, \ldots, g_m$ olsun. Bu fonksiyonların elde edilmesinin ardından, her fonksiyonun hangi noktaları ayırdığını gösteren bir *m*×*m* boyutlu $P_{m\times m} = (p_{il})_{i=1,\ldots,m}^{l=1,\ldots m}$ matrisi, eğer *A* kümesindeki *i*. nokta (a_i,) *l*. fonksiyon (g_l) ile ayrılıyor ise p_{il} =1, diğer durumda p_{il}=0 olacak şekilde, oluşturulsun. Bu aşamada amacımız, en az sayıda fonksiyon kullanarak A kümesindeki mümkün olabilecek en fazla sayıda noktayı B kümesinden ayıracak fonksiyonları seçmektir. Bu problemi çözmek için önerilen iki amaçlı tamsayılı matematiksel programlama yaklaşımı izleyen şekilde verilmiştir.

Karar değişkenleri

x_i=1, eğer a_i noktası seçilen fonksiyonlardan en az biri ile ayrılıyor ise, ve x_i=0 diğer durumda.

$$y_l = \begin{cases} 1, & l.\,\text{fonksiyon seçilirse} \\ 0, & \text{diger durumda} \end{cases}$$

Parametreler

$$p_{il} : \begin{cases} 1, & \text{eger } a_i \text{ noktası } l.\text{fonksiyon ile ayriliyorsa} \\ 0, & \text{diger durumda} \end{cases}$$

Birinci amacımız $f_1(x,y)$, ayırmayı gerçekleştirmek için kullanacağımız fonksiyon sayısının en küçüklenmesi, ikinci amacımız $f_2(x,y)$ ise seçilen fonksiyonlar ile en fazla sayıda noktanın ayrılmasını sağlamak üzere oluşturulmaktadır. Aşağıda, denklem (1.1.2) ile verilen kısıt, l. fonksiyon seçilir ise ve i. nokta l. fonksiyon ile ayrılıyor ise, x_i değişkeninin 1 değerini alabilmesini sağlamaktadır. Bu sayede, eğer i. noktayı ayıran fonksiyonlardan (A kümesinin her noktası birden fazla sayıda fonksiyon ile B kümesinden ayrılabilir) en az bir tanesinin seçilmesi durumunda bu nokta nihai ayırıcı fonksiyon g(x) ile ayrılabilecektir.

$$x_i \le \sum_{l=1}^{m} p_{il} y_l \qquad i = 1, \ldots, m \tag{1.1.2}$$

$$x_i, y_l = \{0,1\} \qquad i, l = 1, \ldots, m \tag{1.1.3}$$

Kısıtları altında

$$f_1(x,y) = enk \sum_{l=1}^{m} y_l \tag{1.1.4}$$

$$f_2(x,y) = enb \sum_{i=1}^{m} x_i \tag{1.1.5}$$

Bu modelin iyi anlaşılabilmesi amacıyla aşağıdaki örneği ele alalım. Tablo 3.1, altı noktadan oluşan bir *A* kümesindeki noktaları ayırmak üzere bulunmuş altı fonksiyonun hangi noktaları ayırdığını göstermektedir. Tablonun satırları noktaları, sütunları ise fonksiyonları göstermektedir. Örneğin ikinci satır, a_2 noktasının g_2 ve g_3 fonksiyonları tarafından ayrıldığını, beşinci sütun ise, g_5 fonksiyonunun a_4 ve a_5 noktalarını ayırdığını göstermektedir. Burada $y_2=y_3=1$ durumu, g_2 ve g_3 fonksiyonlarının seçildiğini göstermekte olup, bu durumda *A* kümesindeki tüm noktaların *B* kümesinden ayrılacağı garantilenmiş olur: $\sum_{i=1} x_i = 6$.

Tablo 3.1 Altı nokta için bir P matrisi

p_{il}	1	2	3	4	5	6
1	1	1	1			
2		1	1			
3			1			
4		1		1	1	
5		1			1	
6		1				1

İki amaçlı tamsayılı matematiksel modelin çözümü ile elde edilecek olan *y* vektörü hangi fonksiyonların son ayırıcı fonksiyonu oluşturacağını bize söyleyecektir. Ayırıcı fonksiyonumuz, seçilen PKF'lerin noktasal en küçüğü olarak aşağıdaki şekilde tanımlanır.

$$\tilde{L} = \{l \in L : y_l = 1\}, \quad g(x) = \min_{l \in \tilde{L}} g_l(x)$$

Burada *L* kümesi, başlangıç adımda oluşturulacak olan ayırıcı polihedral fonksiyonların kümesini göstermektedir.

Başlangıç adımda oluşturulacak fonksiyon sayısı.

Başlangıç adımda oluşturulacak olan ayırıcı fonksiyonlar için merkez noktaları seçilirken tamamen serbest hareket etme imkanı bulunmaktadır. Öyle ki bu noktaları seçerken illa da *A* kümesinin bütün noktalarını seçmek zorunda değiliz. Belki bu noktaların bir kısmı, belki tamamı, belki tamamı artı bazı noktaların konveks örtüsünden (dışbükey birleşiminden) oluşan yeni noktalar, veya *A* kümesinin bazı noktaları için yoğunlaşma merkezi (clustering point) rolünü oynayabilecek noktalar da seçilebilir. Maksat nihai ayırıcı fonksiyonun olabildiğince iyi bir performansa sahip olmasını sağlamaktır.

Eliptik Konik Fonksiyonlar.

Başlangıç adımdaki P_l probleminin çözümü olarak belirlenen ayırıcı polihedral konik fonksiyonlar yerine ***eliptik konik fonksiyonlar*** adını verdiğimiz ve literatürde daha önce hiç görülmemiş olan ve ilk olarak bu proje kapsamında öğrenilecek ve uygulanacak olan fonksiyonlar da kullanılacaktır. Bu fonksiyonların, bir lineer fonksiyonun sadece bir normu ile değil hem de iki normu ile genişletilmesi fikri üzerine oluşturulması düşünülmektedir. Eliptik konik fonksiyonlar aşağıdaki şekilde tanımlanmaktadır:

$$g_{(w,\xi,\sigma,\gamma)} = w'(x-a) + \xi\|x-a\|_1 + \sigma\|x-a\|_2 - \gamma \qquad (1.1.6)$$

İki amaçlı problem sembolik olarak *enk* $\{f_1(x,y); -f_2(x,y)\}$ biçiminde ifade edilebilir. Burada en büyükleme biçiminde verilen ikinci amaç `-' ile çarpılarak enküçükleme haline dönüştürülmüştür.

Bilindiği gibi çok amaçlı problemlerin en yaygın kesin çözüm yöntemlerinden biri ***skalerleştirme*** yöntemleridir. Şimdiye dek birçok skalerleştirme yöntemleri geliştirilmiş ve literatürde tartışılmıştır. Dışbükey olamayan çok amaçlı optimizasyon problemlerinin skalerleştirilmesi için literatürde etkili bir yöntemin bulunmadığını özellikle vurgulamak gerekiyor. Etkili bir yöntem derken, çözülmesi istenen çok amaçlı problemin bütün etkin (efficient) çözümlerinin bulunmasını garantileyen bir yöntem kastedilmektedir. Proje kapsamında yürüttüğümüz çalışmalarımızda bu amaçla yeni bir skalerleştirme tekniğinin kullanılmasını planlıyoruz. Kullanacağımız yöntem, Gasimov tarafından 2001 yılında ispatlanan bir teoreme dayanmaktadır. Şimdiye kadar yürütülen çeşitli çalışmalarda bu teorem kullanılarak çözülmesi zor olarak nitelendirilen ve dışbükey olmayan çok amaçlı problemler için iyi sonuçlar elde edilmiştir. Raporda bu yöntemle ilgili olarak şimdiye dek literatürde bulunmayan ek açıklamalar verilmiştir.

Skalerleştirme

Çok amaçlı eniyileme problemlerinin Pareto (etkin) çözümlerini elde etmek için geleneksel yaklaşımlardan biri de kesin çözümlerin bulunmasını hedefleyen skalerleştirme yöntemidir. Skalerleştirme, aşağıda denklem (1.1.7) 'deki gibi *p* adet amaçtan oluşan çok amaçlı problemleri tek amaçlı optimizasyon problemine "dönüştürmeyi" sağlar. Çeşitli skalerleştirme yöntemleri Ehrgott (2005) tarafından detaylı bir şekilde verilmiştir.

Bu çalışmada, en sık kullanılan skalerleştirme yöntemlerinden biri olan, *Ağırlıklı Toplam Yöntemi* (ATY) ile Gasımov (2001) tarafından önerilen ve çeşitli çalışmalarda (Ozdemir and Gasmov, 2004; Ehrgott vd., 2006; Gasimov vd., (2007) başarı ile uygulanan ve en son haliyle yeni yayınlanmış olan *Konik Skalerleştirme Yöntemi* (KSY) kullanılmıştır, Kasımbeyli (2010).

Konik Skalerleştirme Yöntemi (KSY)

Önce çok amaçlı programlama teorisinden standart tanımları verelim.

$R_+^p = \{y = (y_1,\ldots,y_p): y_k \geq 0,\ k = 1,\ldots,p\}$ ve $Y \subset R^p$ boş olmayan bir küme olsun.

Tanım 1.1.1.

(1) $(\{y\} - R_+^p) \cap Y = \{y\}$ özelliğine sahip olan $y \in Y$ noktasına *Y* kümesinin etkin (veya minimal) noktası denir.

(2) $(\{y\} - \mathrm{int}(R_+^p)) \cap Y = \emptyset$ özelliğine sahip olan $y \in Y$ noktasına *Y* kümesinin zayıf etkin (veya zayıf minimal) noktası denir.

(3) Eğer *y* noktası *Y* kümesinin etkin noktası ve R^p uzayının sıfır noktası da $cl(cone(Y + R_+^p - \{y\}))$ kümesinin etkin noktası ise *y* noktasına *Y* kümesinin (Benson anlamda) has etkin noktası denir; burada *cl(Y)* ile *Y* kümesinin kapanışı, cone(*Y*) ile de *Y* kümesinin konik örtüsü işaretlenmiştir: $cone(Y) = \{\alpha y : y \in Y, \alpha \geq 0\}$.

Aşağıdaki şekilde tanımlanmış

$$\min_{x \in X}\left(f_1(x),\ldots,f_p(x)\right) \tag{1.1.7}$$

çok amaçlı problemi ele alalım. Burada *X* uygun çözümler kümesi, $f_i : X \to R, i = 1,\ldots,p$, ise reel değerli fonksiyonlardır. $f(x) = (f_1(x),\ldots,f_p(x))$ ve $Y = f(X)$ olsun.

Tanım 1.1.2. $x \in X$ elemanı (1.1.7) probleminin bir uygun çözümü, $y = f(x)$ noktası da *Y* kümesinin etkin (zayıf etkin veya has etkin) noktası olsun. Bu durumda *x* elemanına (1.1.7) probleminin etkin çözümü (zayıf etkin çözümü veya has etkin çözümü) denir.

Son yıllarda çok amaçlı problemleri araştırmak için çeşitli interaktif yöntemler ve karar destek sistemleri geliştirilmiştir. Her ne kadar geliştirilmiş olan yöntemlerin büyük çoğunluğu dışbükey (convex) problemler için daha iyi sonuçlar verse de birçok gerçek hayat problemleri maalesef dışbükey olamıyor, nitekim bizim ele aldığımız sınıflandırma problemi de dışbükey değildir.

Skalerleştirme yolu ile çok amaçlı bir problem muhtemelen bir takım parametreler ve/veya ek kısıtlar kullanılarak tek amaçlı bir optimizasyon problemine "indirgenebilir". Kullanılan parametrelerin belli değerleri için tek amaçlı optimizasyon probleminin optimal çözümlerinin, çok amaçlı problemin etkin çözümlerini verdiği bekleniyor. Böyle bir tek amaçlı (skaler) problem, çeşitli skalerleştirici fonksiyonlar kullanılarak oluşturulabilir.

Skalerleştirme yöntemlerinin çoğunda önceden belirlenmiş bir referans noktasına yakın olacak etkin noktaları bulmak amacıyla çeşitli normlar kullanılmıştır. Yu (1973), Zeleny (1973), Steuer (1986), Lewandowski ve Wierzbicki (1989) ve başkaları lp normları kullanmışlar. l∞ normu ve genişletilmiş l∞ normunun, genel olarak sürekli ve kesikli (discrete) çok amaçlı problemlerin etkin çözümlerinin elde edilmesinde faydalı oldukları gözlenmiş ve bu normların kullanılması sonucu iyi bilinen ağırlıklandırılmış Tchebycheff skalerleştirme tekniği ortaya çıkmıştır.

Çeşitli skalerleştirme tekniklerinin kıyaslanması da farklı yollarla yapılmıştır. Wierzbicki (1980, 1998) referans noktaları ile ilgili yöntemler üzerine çalışmaları başlatmış ve çeşitli skalerleştirici fonksiyonların özelliklerini araştırmıştır. Böylece, skalerleştirici fonksiyonların sadece etkin veya Pareto optimal çözümler ürettiği ortaya çıkartılarak bu fonksiyonlarla oluşturulan skalerleştirme yöntemlerinin hedef programlama (goal programming) üzerinde önemli avantaja sahip olduğu kanıtlanmıştır. Son yıllarda iyi bilinen skalerleştirici fonksiyonlar ve onların çeşitli modifikasyonları Miettinen ve Makela (2002) tarafından kıyaslanmıştır.

Farklı skalerleştirme tekniklerinin özellikleri araştırılarak bu tekniklerin kalite ve performansları ile ilgili aşağıdaki sorular geliştirilebilir:

- Bu teknik her zaman etkin çözümler üretiyor mu?
- Eğer böyle ise bütün etkin çözümleri yakalamak mümkün müdür?
- Bu teknik, karar vericinin farklı amaçlarla ilgili tercihlerini (ağırlık katsayılarını) ve karar vericinin etkin çözümlerle ilgili tercihlerini (referans noktalarını) göz önünde bulunduruyor mu?

Yukarıdaki ikinci soru, yani bir tekniğin bütün etkin çözümleri yakalayıp yakalamayacağı sorusu, özellikle karar vericinin tercihlerine cevap verecek etkin çözümlerin bulunup bulunamayacağı ile ilgili olup, bir skalerleştirme tekniğinden beklenen en önemli özelliklerden ve de cevap verilmesi en zor sorulardan biridir.

Farklı skalerleştirme teknikleri ile ilgili yapılan basit bir araştırma hiç de bütün yöntemlerin buradaki soruların tamamına pozitif cevap veremediğini göstermektedir.

Bu projede biz, ana fikri, Benson has etkin noktaların karakterizasyonu için Gasimov (2001) tarafından verilmiş bir teoreme dayanan bir skalerleştirme tekniğini kullanacağız. Gasimov, monoton artan ve dışbükey fonksiyonların özel bir sınıfını tanımlamış ve bu sınıfa ait fonksiyonları kullanarak problem üzerine hiç bir kısıtlayıcı koşullar (dışbükeylik ve sınırlılık gibi) kullanmaksızın Benson has etkin çözümler için önemli karakterizasyon teoremleri ispatlamıştır.

Bu teoremler kullanılarak geliştirilen konik skalerleştirme yönteminin (KSY) dayandığı temel fikir, Pareto etkin çözümlerin bulunmasında destek konilerinin kullanılmasıdır. Yöntemin en büyük üstünlüğü de çok geniş bir problem sınıfına hitap etmesidir. KSY, daha önce de birçok dışbükey olmayan çok amaçlı programlama problemine uygulanmış ve başarılı sonuçlar elde edilmiştir. Özdemir ve Gasimov, KSY'yi çok amaçlı fakülte ders atama problemine uygulamışlardır (Özdemir and Gasimov, 2004). Ayrıca Gasimov vd. (2007), KSY'yi 1.5 boyutlu kesme problemlerine uygulamışlar ve elde ettikleri sonuçları Ağırlıklı Toplam Yöntemi ile karşılaştırmışlardır.

Aşağıda Gasimov tarafından ispatlanmış karakterizasyon teoremlerinden birini ispatsız olarak veriyoruz (bakınız, Theorem 1, Gasimov, 2001).

$$W = \{(w,\alpha) \in R_+^p \times R_+ : 0 \le \alpha < \min\{w_1,\dots,w_p\}\} \tag{1.1.8}$$

olsun.

Teorem 1.1.1. (Teorem 1; Gasimov, 2001) $x \in X$ elemanının, belli bir $(\alpha, w) \in W$ için izleyen skaler enküçükleme probleminin en iyi çözümü olduğunu varsayalım.

$$\underset{x\in X}{enk} \sum_{i=1}^{p} w_i f_i(x) + \alpha \sum_{i=1}^{p} |f_i(x)| \tag{1.1.9}$$

O zaman $x \in X$ elemanı, (1.1.7) probleminin bir Benson has etkin çözümüdür.

Bu teoremin tersi Gasimov tarafından sadece, uzay Bishop-Phelps konisi adı verilen özel tip konilerle sıralandığı durum için ispatlanmıştır. Asıl merak konusu olan soru, acaba her Benson has etkin nokta için bu noktayı (1.1.9) probleminin optimal çözümü olarak elde edebileceğimiz bir $(w,\alpha) \in W$ skalerleştirme parametresi çifti bulabilir miyiz? Bu soruya cevap bulabilmek için Teorem 1.1.1 de (1.1.8) ile tanımlanan skalerleştirme parametreleri kümesinden alınan her $(w,\alpha) \in W$ skalerleştirme çifti için $g_{(w,\alpha)} : R^p \to R$ olarak kullanılan ve

$$g_{(w,\alpha)}(y) = \sum_{i=1}^{p} w_i y_i + \alpha \sum_{i=1}^{p} |y_i| \tag{1.1.10}$$

veya

$$g_{(w,\alpha)}(y) = w' y + \alpha \|y\|_1 \tag{1.1.11}$$

şeklinde tanımlanan skalerleştirici fonksiyonlar ailesinin özelliklerini açıklayan aşağıdaki teoremlerle yola çıkıyoruz.

Teorem 1.1.2. Aşağıdakiler doğrudur:

(i) $\alpha > 0$ olmakla her $(w,\alpha)\in W$, çifti için (1.1.11) şeklinde tanımlanan $g_{(w,\alpha)}$ skalerleştirici fonksiyonunun alt seviye kümesi

$$C^{-}(w,\alpha)=\{y\in R^{p} : w^{'}y+\alpha\|y\|_{1}\le 0\} \tag{1.1.12}$$

dışbükey, kapalı, pointed bir koni olup, $-R_{+}^{p}$'yi içeriyor; C konisi, $C\cap(-C)=\{0\}$ koşulunu sağlarsa pointed adlanır.

(ii) $(w,\alpha_1)\in W$, $(w,\alpha_2)\in W$ ve $\alpha_1\le\alpha_2 \Rightarrow C^{-}(w,\alpha_2)\subseteq C^{-}(w,\alpha_1)$.

İspat. (i) $\alpha > 0$ olacak şekilde $(w,\alpha)\in W$ olsun. $C^{-}(w,\alpha)$'nın bir koni olduğu açıktır. Öte yandan $g_{(w,\alpha)}$ fonksiyonu dışbükey ve sürekli bir fonksiyon olduğundan, bu fonksiyonun alt seviye kümesi olan $C^{-}(w,\alpha)$ da dışbükey ve kapalı bir konidir. $y\in C^{-}(w,\alpha)$ alalım. Bu durumda

$$w^{'}y+\alpha\|y\|_{1}\le 0 \tag{1.1.13}$$

$-y\in C^{-}(w,\alpha)$ olduğunu varsayarsak,

$$-w^{'}y+\alpha\|-y\|_{1}\le 0 \tag{1.1.14}$$

elde ederiz. Bu durumda (1.1.13) ve (1.1.14) eşitsizliklerinden $\alpha > 0$ da dikkate alınırsa y=0 elde edilir ki, bu da $C^{-}(w,\alpha)$ 'nin bir pointed koni olduğu anlamına gelir.

Şimdi $y\in -R_{+}^{p}$ olsun. O zaman $-y\in R_{+}^{p}$ veya bir başka şekilde $-y\ge 0$. Bu durumda $(w,\alpha)\in W$ olduğundan $w^{'}(-y)-\alpha\|-y\|_{1}\ge 0$ veya sağ tarafa geçirirsek $w^{'}y+\alpha\|y\|_{1}\le 0$ elde ederiz ki, bu da $C^{-}(w,\alpha)$ nin tanımına göre $y\in C^{-}(w,\alpha)$ anlamına gelmektedir. Böylece (i) ispatlanmış oldu.

(ii) Şimdi $(w,\alpha_1)\in W$, $(w,\alpha_2)\in W$ ve $\alpha_1\le\alpha_2$ ve $y\in C^{-}(w,\alpha_2)$ olsun. Buradan kolayca

$$0\ge w^{'}y+\alpha_2\|y\|_{1}\ge w^{'}y+\alpha_1\|y\|_{1}$$

Elde ederiz ki bu da $y\in C^{-}(w,\alpha_1)$ anlamına gelir ve böylece teoremin de ispatı tamamlanmış bulunuyor.

Şimdi bu teoremler kullanılarak (1.1.7) çok amaçlı probleminin nasıl skalerleştirildiğini açıklayalım.

Varsayalım ki karar verici, $w_i\ge 0,\ w_1+...+w_p=1$ olacak şekilde bir $w=(w_1,...,w_p)$ ağırlık vektörü ve $a=(a_1,...,a_p)$ şeklinde bir referans noktası vererek kendi tercihlerini belirtiyor.

Hatırlatalım ki referans noktası, karar vericinin bu noktaya "yakın" bir çözüm elde edilmesini istediği durumlarda kullanılabilir.

Gayet açıktır ki, eğer $\hat{x} \in X$, (1.1.7) probleminin bir Pareto etkin çözümü ise izleyen "kaydırılmış problem"in de Pareto etkin çözümüdür.

$$\min_{x \in X} \left(f_1(x) - a_1, \ldots, f_p(x) - a_p \right) \tag{1.1.15}$$

Bu durumda (1.1.9) a benzer bir skaler problem yazabiliriz:

$$\underset{x \in X}{enk} \sum_{i=1}^{p} w_i \left(f_i(x) - a_i \right) + \alpha \sum_{i=1}^{p} \left| f_i(x) - a_i \right| \tag{1.1.16}$$

Burada $w = (w_1, \ldots, w_p)$ vektörü, $w_i \geq 0,\ w_1 + \ldots + w_p = 1$ şartlarını sağlayan ağırlık vektörü ve $a = (a_1, \ldots, a_p)$ vektörü de referans noktasıdır, α sayısı genişletme parametresi olup $0 \leq \alpha < \min\{w_1, \ldots, w_p\}$ koşulunu sağlamaktadır.

Teorem 1.1.1 diyor ki, (1.1.16) skaler probleminin her bir optimal çözümü (1.1.15) çok amaçlı probleminin has etkin çözümüdür ve dolayısı ile de (1.1.7) probleminin has etkin çözümüdür. α=0 durumunda (1.1.16) skaler probleminin ağırlıklandırılmış toplam yönteminin amaç fonksiyonuna dönüştüğü açıkça görünmektedir.

3.2.2 Ağırlıklandırılmış toplam yöntemi

Bu yönteme göre amaç fonksiyonları, pozitif w_i ağırlıkları ile çarpılıp toplanarak problem tek amaçlı optimizasyon problemi haline dönüştürülüyor.

$$\underset{x \in \chi}{enk} \sum_{i=1}^{p} w_i f_i(x) \tag{1.1.17}$$

Teorem 1.1.3. (Geoffrion, 1968) $i = 1, \ldots, p$ için $w_i > 0$, $\sum_{i=1}^{p} w_i = 1$ şartını sağlayan pozitif ağırlıklar olsun. Eğer $\hat{x}$, (1.1.17) probleminin eniyi çözümü ise o zaman (1.1.7) probleminin de has etkin çözümüdür.

Sınıflandırma problemi için geliştirdiğimiz tamsayılı modeldeki $f_1(x, y)$ ve $f_2(x, y)$ amaç fonksiyonları sırasıyla w_1 ve w_2 ağırlıkları ile çarpılarak birleştirilir aşağıdaki gibi bir skaler amaç fonksiyonu elde edilir.

$$\underset{x \in X}{enk}\ f(x, y) = w_1 \sum_{l} y_l - w_2 \sum_{i} x_i \tag{1.1.18}$$

burada ikinci amaç fonksiyonunun ağırlıklı toplamda eksi işareti ile bulunması, orijinal problemde bu amaç fonksiyonunun enbüyükleme (maksimizasyon) şeklinde olmasından kaynaklanmaktadır.

(1.1.17) problemi $a = (a_1, \ldots, a_p)$ referans noktası da dikkate alınarak yazılırsa, bu problemin her hangi bir $\bar{x}$ çözümü için aşağıdaki münasebet yazılır:

$$w\left(f(x) - a \right) \geq w\left(f(\bar{x}) - a \right) \quad \forall x \in X \ .$$

$y = f(x), Y = f(X)$ işaretlemelerini dikkate alınarak bunu aşağıdaki gibi de yazabiliriz:

$$w(y - a) \geq w(\bar{y} - a) \quad \forall y \in Y \ ,$$

veya

$$w(y-\bar{y})\geq 0 \quad \forall y\in Y.$$

Bu ise, $\bar{y}$ noktasının *Y* kümesinin $H=\{y\in R^p : w(y-\bar{y})=0\}$ hiperdüzlemine nazaran destek noktası olduğu anlamına gelmektedir.

Şimdi α sayısı $0\leq \alpha < \min\{w_1,...,w_p\}$ koşulunu sağlayacak şekilde seçilmiş olsun ve de $\bar{x}$ noktası da (1.1.16) probleminin bir optimal çözümü olsun. Bu durumda

$$w(f(x)-a)+\alpha\|f(x)-a\|_1 \geq w(f(\bar{x})-a)+\alpha\|f(\bar{x})-a\|_1 \quad \forall x\in X,$$

veya

$$w(y-a)+\alpha\|y-a\|_1 \geq w(\bar{y}-a)+\alpha\|\bar{y}-a\|_1 \quad \forall y\in Y=f(X).$$

Son eşitsizlik aşağıdaki gibi de yazılabilir:

$$w(y-\bar{y})+\alpha\|y-\bar{y}\|_1 \geq 0 \quad \forall y\in Y \tag{1.1.19}$$

(1.1.19) ilişkisi aşağıdaki gibi yorumlanabilir. Tepe noktası $\bar{y}$ noktasında bulunan

$$C(w,\alpha;\bar{y})=\{y\in R^p : w(y-\bar{y})+\alpha\|y-\bar{y}\|_1 = 0\} \tag{1.1.20}$$

şeklinde tanımlanmış olan $C(w,\alpha;\bar{y})$ konik yüzeyi *Y* kümesine $\bar{y}$ etkin noktasında aşağıdaki anlamda destek konisidir:

$$Y\subset C^+(w,\alpha;\bar{y})=\{y\in R^p : w(y-\bar{y})+\alpha\|y-\bar{y}\|_1 \geq 0\}$$

ve

$$\bar{y}\in Y\cap C(w,\alpha;\bar{y})$$

Son iki münasebetle tarif edilen destekleme felsefesi eşdeğer olarak aşağıdaki gibi de yazılabilir:

$$Y-\{\bar{y}\}\subset C^+(w,\alpha)=\{y\in R^p : wy+\alpha\|y\|_1 \geq 0\}$$

ve

$$\{0\}\subset(Y-\{\bar{y}\})\subset C(w,\alpha)=\{y\in R^p : wy+\alpha\|y\|_1 = 0\}$$

Yukarıda analiz edildiği gibi, $\alpha=0$ durumunda $C(w,\alpha)$ konisi hiperdüzleme dönüşüyor. Teorem 1.1.2 ye göre bu koni, α arttıkça küçülüyor ve α 'nın mümkün olabilecek en büyük değerinde (yani, $\alpha=\min\{w_1,...,w_p\}$) verilmiş *w* ağırlık vektörü için en dar koniye dönüşüyor. Böylece verilmiş *w* ağırlık vektörü için α sayısını kendi değer aralığında sürekli olarak değiştirerek (1.1.16) problemi çözülerek, karar vericinin tercihlerine cevap verecek tüm etkin noktaların bulunabileceği görülmektedir.

Sınıflandırma probleminin çözümü için önerilen iki amaçlı tamsayılı matematiksel modelin çözümünde, KSY için kullanılan parametreler ve amaç fonksiyonu izleyen şekilde verilmiştir.

Parametreler :

c_1 : birinci amaç için referans noktası

c_2 : ikinci amaç için referans noktası
α : genişletme parametresi, 0 < $\alpha \leq$ enk {w_1, w_2}

Amaç Fonksiyonu:

$$enk\ f(x,y) = w_1\left(\sum_k y_k - c_1\right) + w_2\left(-\sum_i x_i - c_2\right) + \alpha\left(\left|\sum_k y_k - c_1\right| + \left|-\sum_i x_i - c_2\right|\right) \quad (1.1.21)$$

3.2.3 Tamsayılı modelin geliştirilmesi

İki amaçlı tamsayılı matematiksel model temelli algoritma, P matrisine farklı merkez noktaları için elde edilmiş yeni fonksiyonlara ait sütunların eklenmesiyle genişletilmiştir. Bu amaçla bir çok farklı strateji geliştirme mümkün olmakla beraber, bu aşamada sadece A kümesinden seçilen merkez noktalarına göre oluşturulmuş en yüksek başarıyı veren belirli sayıda fonksiyondan yararlanılmıştır. Bu fonksiyonların ayırdıkları noktaların ortalaması alınarak elde edilen nokta merkez kabul edilmiş ve yeni bir fonksiyon bulunmuştur. Ayrıca polihedral konik fonksiyonların elde edilmesinde kullanılan doğrusal modelde, B kümesinden ayrılacak noktalar için bir tolerans tanımlanmış ve farklı tolerans değerleri için problemler çözülerek eğitim ve test başarılarının birbirlerine yaklaştığı yani genelleştirme başarısının arttığı gösterilmiştir.

İki amaçlı tamsayılı modelde "en az sayıda fonksiyon seçilmesi" amacı kısıt olarak değiştirilerek model tek amaçlı hale getirilmiş ve farklı sayıda fonksiyon seçilmesi ile elde edilen ayırıcı fonksiyonların başarıları üzerine detaylı analizler yapılmıştır.

$$x_i \leq \sum_{l=1}^{m} p_{il} y_l \quad i = 1,\ldots,m \quad (1.1.2)$$

$$x_i, y_l = \{0,1\} \quad i,l = 1,\ldots,m \quad (1.1.3)$$

$$f_2(x,y) = \sum_{i=1}^{m} x_i \leq S \quad (1.1.4)$$

Kısıtları altında

$$enk\ f_1(x,y) = \sum_{l=1}^{m} y_l \quad (1.1.5)$$

3.2.4 Sayısal denemeler

İki amaçlı tamsayılı yaklaşımın başarımını göstermek için literatürden alınan Liver, WBCP, Hearth, Ionosphere ve Diabets veri kümeleri üzerinde denemeler yapılmıştır.

Heart veri kümesi için w_1=30, w_2 = 20 ağılıkları kullanılarak elde edilen sonuçlar Tablo 3.2 ile sunulmaktadır. Bu tablodaki ilk üç sütunda α, c_1 ve c_2 parametreleri, sonraki sütunlarda ise problemin çözümü sonucunda elde edilen değerler gösterilmektedir. Heart veri kümesi için en iyi test başarımı 27. deneyde elde edilmiştir.

Tablo 3.2 w_1=30, w_2=20 ağılıkları için Heart veri kümesi kullanılarak elde edilen on kez çapraz doğrulama sonuçları

Deney	α	c_1	c_2	$\overline{\sum y_l}$	$\overline{\sum x_i}$	**Eğitim**	**Test**

1	18	9	-110	9.7	-108.4	0.9443	0.791
2			-109	9.2	-108	0.9356	0.787
3			-108	9.2	-107.4	0.9405	0.774
4			-107	8.9	-106.9	0.9386	0.784
5		10	-110	10.2	-109	0.9465	0.791
6			-109	9.9	-108.4	0.9443	0.774
7			-108	9.5	-107.7	0.9416	0.781
8			-107	9.2	-107.3	0.9401	0.787
9		11	-110	10.6	-109.4	0.948	0.771
10			-109	10.3	-108.9	0.9461	0.771
11			-108	9.9	-108.3	0.9439	0.787
12			-107	9.4	-107.7	0.9416	0.794
13	19	9	-110	9.7	-108.4	0.9443	0.774
14			-109	9.5	-108	0.9428	0.784
15			-108	9.2	-107.4	0.9405	0.774
16			-107	8.9	-106.9	0.9386	0.787
17		10	-110	10.2	-109	0.9465	0.791
18			-109	9.9	-108.4	0.9443	0.784
19			-108	9.5	-107.7	0.9416	0.791
20			-107	9.2	-107.3	0.9401	0.794
21		11	-110	10.6	-109.4	0.948	0.774
22			-109	10.3	-108.9	0.9461	0.771
23			-108	9.9	-108.3	0.9439	0.794
24			-107	9.2	-107.3	0.9401	0.791
25	20	9	-110	9.7	-108.4	0.9443	0.781
26			-109	9.5	-108	0.9428	0.787
27			**-108**	**9.2**	**-107.4**	**0.9405**	**0.798**
28			-107	8.9	-106.7	0.9383	0.787
29		10	-110	10.2	-108.9	0.9461	0.787
30			-109	9.9	-108.4	0.9443	0.787
31			-108	9.5	-107.6	0.9416	0.784
32			-107	9.2	-107	0.9401	0.781
33		11	-110	10.6	-109.3	0.9476	0.784
34			-109	10.3	-108.8	0.9458	0.774
35			-108	9.9	-108.1	0.9435	0.791
36			-107	9.2	-107.1	0.9401	0.781

Her veri kümesi üzerinde Heart veri kümesi için yapılan deneylere benzer deneyler yapılarak elde edilen en iyi test başarımı sonuçları Tablo 3.3 ile verilmiştir.

Tablo 3.3 On kez çapraz doğrulama sonuçları

Problem	w_1	w_2	α	c_1	c_2	$\overline{\sum y_l}$	$\overline{\sum x_i}$	**Eğitim**	**Test**
Liver	36	14	10	10	-150	9.6	-152.1	79.07	70.14
WBCP	10	40	10	12	-130	10.2	-133.2	79.95	70.62
Heart	30	20	20	9	-108	9.2	-107.4	94.05	79.8
Ionosphere	30	20	20	3	-112	1.7	-111.6	99.43	90.03
Diabets	28	22	15	40	-162	40.8	-160.8	88.37	73.31

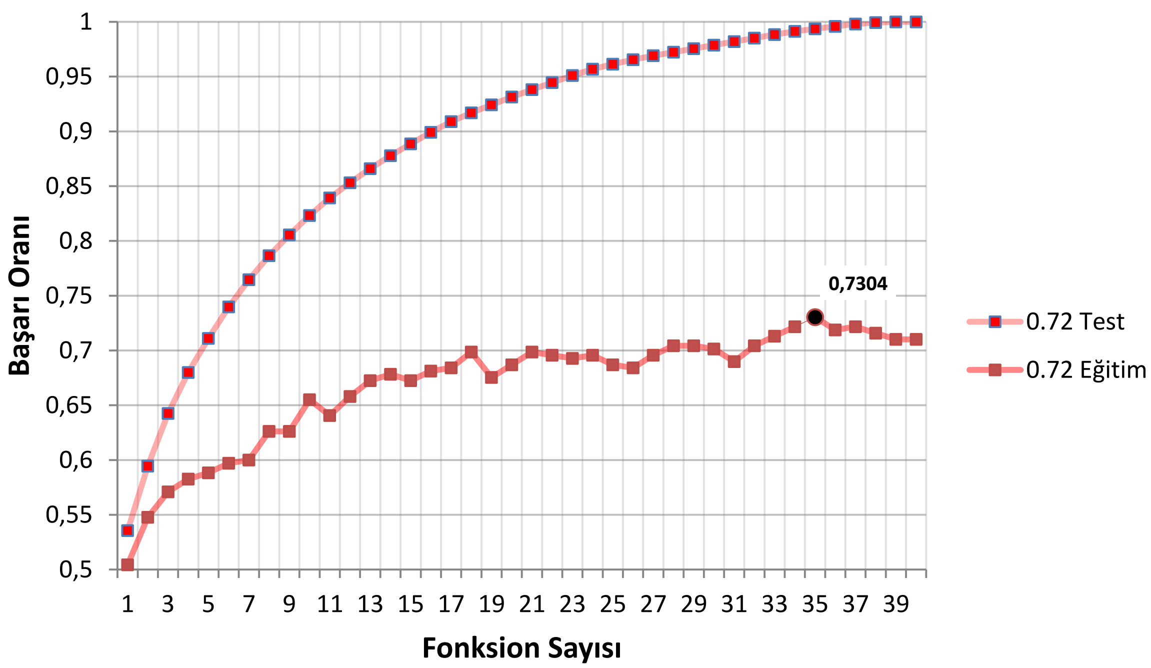
Önerilen iki amaçlı tamsayılı matematiksel programlama yaklaşımı sezgisel bir yaklaşımdır. Bu yaklaşımda önce m adet PKF oluşturulmakta ve bu fonksiyonlar yardımıyla tamsayılı programlama probleminin p_{il} parametreleri hesaplanmaktadır. Daha sonra tamsayılı problem konik skalerleştirme yaklaşımı ile skalerleştirerek çözülmektedir. Tablo 3.3 ile verilen beş veri kümesi için farklı amaç ağırlıkları ve diğer konik skalerleştirme parametrelerine göre birçok test yapılmıştır. Heart, Diabet, Ionosphere veri kümeleri için daha önce geliştirilmiş olan PKF algoritması ile elde edilen sonuçlardan daha iyi test başarımı sonuçları elde edilmiştir. Bunun yanında, PKF algoritmasına göre eğitim ve test başarımları arasındaki fark azalmış, yani genelleştirme başarımı artmıştır.

Tek amaçlı tamsayılı algoritma farklı

Tablo 3.4 Liver Problemi için tek amaçlı tamsayılı modelin farklı tolerans değerlerine göre on kez çapraz doğrulama sonuçları

Tolerans	fonk Say.	$\sum y_l$	Egitim_B	Test_B
0.6	48	45.9	1	0.7217
0.65	40	39.8	0.9977	0.7217
0.7	31	31	0.9771	0.7246
0.72	35	35	0.9936	**0.7304**
0.8	32	32	0.9968	0.7217

Liver probleminin 0.72 tolerans değerine göre fonksiyon sayısı - eğitim ve test başarı oranları grafiği aşağıda verilmiştir. Bu problem için grafikte fonksiyon sayısı arttıkça eğitim başarısının arttığı Şekil 3.1 ile gösterilmektedir.



Şekil 3.1 Liver problemi için on kez çapraz doğrulama sonuçları

Tablo 3.5 Tek amaçlı matematiksel model için on kez çapraz doğrulama sonuçları

Problem	Tolerans	Fonksiyon Sayısı	Eğitim	Test
Liver	0.72	35	99.36	73.04
WPBC	1.7	1	76.58	75.77

Heart	0	10	95.81	81.48
Diabets	1.05	4	76.75	76.17

3.3 Çok sınıflı sınıflandırma algoritmalarının geliştirilmesi

Sınıflandırma problemi denildiğinde, çoğunlukla iki sınıflı problemler akla gelmektedir. İki sınıflı problemlerin önemi, birçok alanda çok sayıda gerçek hayat problemi olarak karşımıza çıkması ve bu problemler için geliştirilen yöntemlerin çok-sınıflı problemlere de uygulanabilmesidir. Sınıf sayısı ikiden büyük olduğunda, sınıflandırma problemleri çok-sınıflı olarak adlandırılır. Optik karakter tanıma, ses tanıma ve bir kelimenin cümledeki görevinin belirlenmesi gibi problemler çok-sınıflı problemlere örnek olarak gösterilebilir.

Literatürde, çok-sınıflı problemler için birçok algoritma geliştirildiği halde herhangi bir yaklaşımın diğerlerine tam olarak üstünlük sağladığı iddia edilememektedir. Ayrıca ikili sınıflandırma yaklaşımlarının etkin bir şekilde çoklu sınıflandırma yaklaşımlarına genişletilmesinin hala güncelliğini koruyan bir araştırma alanı olduğuna dikkat çekilmektedir (Hsu ve Lin, 2002).

İki sınıflı problemlerin çözümü için geliştirilen PKF algoritması ile iki amaçlı tamsayılı programlama yaklaşımının (İTS), çok-sınıflı problemlerin çözümünde nasıl kullanıldığı izleyen paragraflarda açıklanmaktadır. Çok-sınıflı sınıflandırma probleminin çözümü genellikle ikili sınıflandırma yöntemleri kullanılarak gerçekleştirilmektedir. Bu yaklaşım, problemin iki sınıflı alt problemlere ayrıştırılması ve ikili sınıflandırma yöntemleri ile elde edilen sınıflandırıcılar kullanılarak çoklu sınıflandırıcının oluşturulması biçiminde uygulanmaktadır.

İkili sınıflandırıcılar kullanılarak çok-sınıflı sınıflandırıcıların oluşturulmasında yaygın kullanılan iki yaklaşım bulunmaktadır. Bu yaklaşımlardan biri, her sınıf için bir sınıflandırıcı oluşturarak, ilgili sınıfa ait örnekleri, veri kümesinin kalan diğer örneklerinden ayıran, *bire karşı hepsi* (*1eh*) yaklaşımıdır. Diğer yaklaşım ise, tüm sınıf ikilileri için sadece bir sınıflandırıcının oluşturulduğu *bire karşı bir* (*1e1*) yaklaşımıdır. Bunlara ek olarak, literatürde yönlü çevrimsiz serim (*YCS*) gibi yaklaşımlar da bulunmaktadır. Bire karşı bir ve bire karşı hepsi yaklaşımlarının özellikleri ve PKF algoritması ve iki amaçlı tamsayılı programlama yaklaşımı ile nasıl kullanıldıkları aşağıdaki paragraflarda açıklanmıştır.

3.3.1 Bire karşı bir (1e1)

Çiftli sınıflandırma olarak da adlandırılan bu yaklaşımda olası her sınıf çifti için bir sınıflandırıcı oluşturulur. Sınıf sayısının *d* olduğu bir sınıflandırma probleminde, d(d-1)/2 ikili sınıflandırıcı ortaya çıkar. Sınıflandırıcı sayısı sınıf sayısına bağlı olarak artacağından, sınıf sayısı arttıkça problemlerin bu yaklaşımla ele alınması daha zor hale gelir. Örneğin, dört sınıflı bir problem için altı tane ikili sınıflandırıcı ortaya çıkarken, sınıf sayısı 10 olduğunda ikili sınıflandırıcıların sayısı 45 olacaktır. Dolayısıyla hem eğitim hem de test için daha fazla çaba harcamak gerekecektir.

1e1 yaklaşımının her bir sınıflandırıcısında sadece iki sınıfa ait örnekler kullanıldığından, sınıflandırıcıyı oluşturmak için ele alınan problemin boyutu *1eh* yaklaşımına göre daha küçük olacaktır. Bu durumda daha az üst üste gelen sınıf olacağından ele alınan problemler genellikle daha kolay olmaktadır (Schaolkopf ve Smola,, 2002). Örneğin, her biri 10 örnekten oluşan 10 sınıflı bir problem düşünelim. Bu problemde oluşturulacak 45 sınıflandırıcının her biri sadece 20 örnek ile oluşturulacaktır.

1e1 yaklaşımda ikili sınıflandırıcıların tümü elde edildikten sonra test aşamasına geçilir. Herhangi bir test noktasının hangi sınıfa ait olduğunu belirlemek üzere farklı yaklaşımlar önerilmiştir. Bu amaçla, genellikle Friedman (1996) tarafından önerilen ``oylama'' yada diğer bir ismi ile ``en çok kazanan'' stratejisi kullanılmaktadır (Hsu ve Lin, 2002, Ou vd. 2004).

Yeni bir test noktasının oylama yöntemi ile hangi sınıfa ait olacağı belirlenirken, bu nokta tüm sınıflandırıcılarda sınanır. Test noktası, ilgili sınıflandırıcı tarafından hangi sınıfa atandı ise o sınıfın oyu bir artar. Tüm sınıflandırıcılar ile oylama yapıldıktan sonra, test noktası en çok oyu alan sınıfa atanır. Bu oylamada en fazla oyu alan birden çok sınıf olması durumunda, atamanın hangi sınıfa yapılacağını belirlemek için farlı öneriler ortaya atılmıştır. Ou vd. (2004), bu eşitliğin bozulabilmesi için önsel olasılıkların veya oylar için belirli ağırlıkların kullanılabileceğini belirtmiştir. Hsu ve Lin (2002) ise eşit oy alan sınıflardan en küçük sınıf numarasına sahip olana atama yapmıştır. Benzer durumda Krebel (1999) ise en yakın sınıfa atama kuralını uygulamıştır. Şekil 3.2 ile *1e1* yaklaşımı için sözde kod verilmektedir.

- Eğitim ve test kümelerini oluştur
- Eğitim kümesindeki d sınıfa ait noktaları kullanarak $d(d-1)/2$ ayırıcı fonksiyon oluştur
- A kümesindeki tüm elemanlar için $\forall a_i \in A$, sınıf atamalarını belirle
- en çok oy alan sınıfı belirle
- En çok sayıda oyu alan birden çok sınıf var ise en küçük fonksiyon değerine sahip sınıfa atama yap
- Eğitim ve test kümelerindeki noktalar için hatalı sınıflandırma matrislerini oluştur
- k. Çapraz doğrulama için eğitim başarısını hesapla
- Ortalama eğitim başarısını hesapla
- Ortalama test başarısını hesapla

Şekil 3.2. 1e1 yaklaşımı için sözde kod.

3.3.2 Bire karşı hepsi (1eh)

Bu yaklaşımda her sınıf için bir tane olmak üzere, sınıf sayısı kadar sınıflandırıcı oluşturulur. Her bir sınıflandırıcı, bir sınıfın noktalarını geri kalan sınıflardaki noktalardan ayıracak şekilde elde edilir.

Yeni bir test noktasının hangi sınıfa ait olduğunun belirlenmesinde üç farklı durum ile karşılaşılmaktadır. Test noktası, sınıflandırıcılardan sadece biri ile sınıflandırılıyor ise bu durumda atama ilgili sınıfa yapılır. Eğer nokta birden çok sınıflandırıcı ile sınıflandırılıyor ise ya da hiç bir sınıflandırıcı ile sınıflandırılamıyor ise atama doğrudan yapılamaz. Benzer durumlarla karşılaşıldığında, Ou vd. (2004) farklı karar kurallarının kullanılabileceğini belirtmektedir. Şekil 3.3 ile *1eh* yaklaşımı için sözde kod verilmektedir.

```
• Eğitim ve test kümelerini oluştur
• Eğitim kümesindeki d sınıf için d ayırıcı fonksiyon
  oluştur
• A kümesindeki tüm elemanlar için ∀a_i ∈A , sınıf
  atamalarını belirle
• en küçük fonksiyon değerine sahip sınıfı ata
• Eğitim ve test kümlerindeki noktalar için hatalı
  sınıflandırma matrislerini oluştur
• k. Çapraz doğrulama için eğitim başarısını hesapla
• Ortalama eğitim başarısını hesapla
• Ortalama test başarısını hesapla
```

Şekil 3.3 1*eh* yaklaşımı için sözde kod

3.3.3 Doğrusal fonksiyonların farklı normlar ile genişletilmesi

Konik fonksiyonların sıfır noktasındaki seviye kümeleri, yani $\{x : g(x)=0\}$ şeklinde tanımlanan kümeler, uzayı, bu seviye kümesinin "altında" ($\{x : g(x)\leq 0\}$) ve "üstünde" ($\{x : g(x)\geq 0\}$) kalan kısımlar ("yarı-uzaylar") olmak üzere iki bölgeye ayırmaktadır. Bu sebeple konik fonksiyonlar kullanılarak elde edilen ayırıcı yüzey, bu fonksiyonların seviye kümelerinin yardımıyla oluşturulmaktadır. Farklı normların kullanılması ayırma yetenekleri birbirinden farklı olan çeşitli ayırıcı yüzeylerin elde edilmesini sağlamaktadır. Bu sebeple araştırmamızın bir boyutu da daha yüksek sınıflandırma başarıları elde etmek amacıyla farklı ayırıcı yüzeyler ile sonuçlanacak l_2 ve l_{max} normlarının kullanımıdır. l_1 ve l_{max} normlarının seviye kümeleri farklı dışbükey polihedral kümeler şeklinde ortaya çıkmaktadır. Diğer taraftan l_2 normu konik elipsoid şeklinde ayırıcı yüzeylerin elde edilmesine imkân sağlamaktadır.

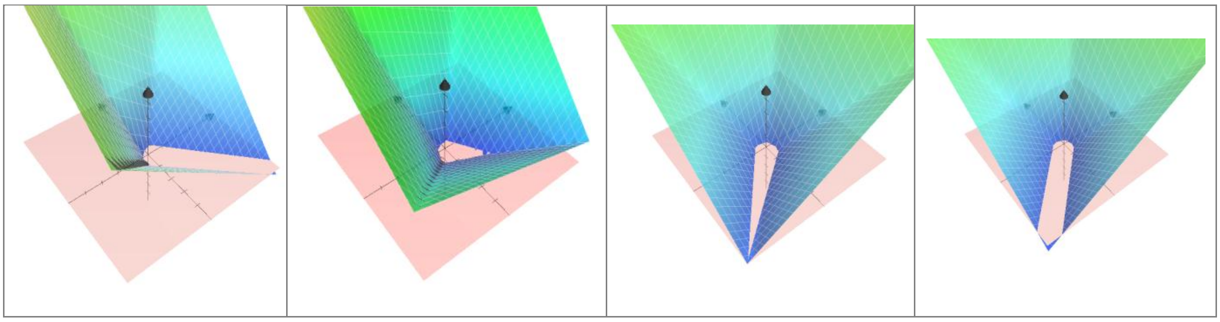
Bu projede geliştirdiğimiz çeşitli sınıflandırma algoritmalarının temelinde duran bu fonksiyonların sağladığı üstün performansı anlamak için bu fonksiyonların grafiklerinin, onları oluşturan parametrelere bağlı olarak nasıl değiştiğini geometrik olarak yorumlayalım.

l_1 normu : Aşağıda denklem (1.2.1) ile verilen ve l_1 normu ile oluşturulan g_{l1} fonksiyonu bir polihedral konik fonksiyonu ifade etmektedir. R^2'de tanımlanan ve aşağıda denklem (1.2.2) ile sunulan bir örnek için fonksiyonun grafiğinin ve seviye kümesinin w_1, w_2, ζ ve γ parametrelerine göre değişimi sırasıyla Şekil 3.4, Şekil 3.5, Şekil 3.6 ve Şekil 3.7 ile gösterilmektedir.

$$g_{l1}(x) = wx + \xi\|x\|_1 - \gamma \quad (1.2.1)$$

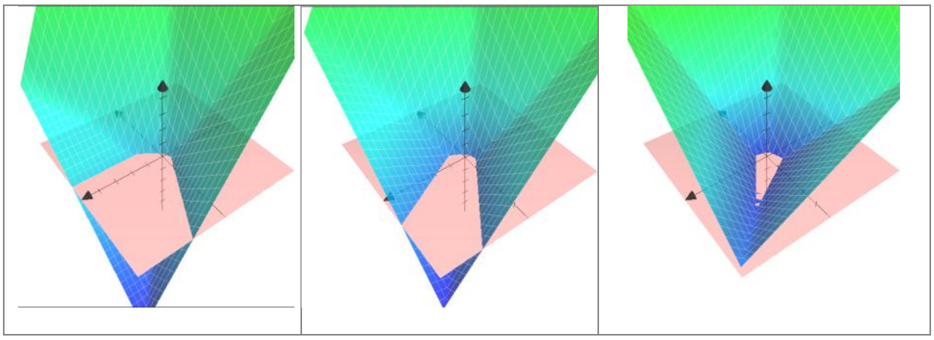
$$g_{w,\xi,\gamma}(x_1,x_2) = w_1(x_1 - a_{01}) + w_2(x_2 - a_{02}) + \xi(|x_1 - a_{01}| + |x_1 - a_{02}|) - \gamma \quad (1.2.2)$$

Merkez noktası olarak a_0=(-2,3) noktası ve w_2 = 0.1, ξ = 1.3 ve = 2 olarak belirlendiği $g(x)=w_1(x_1+2)+0.1(x_2-3)+1.3(|x_1+2|+|x_2-3|) - 2$ fonksiyonu için w_1 parametresinin [-1,1] aralığındaki değişimi Şekil 3.4 ile gösterilmektedir.



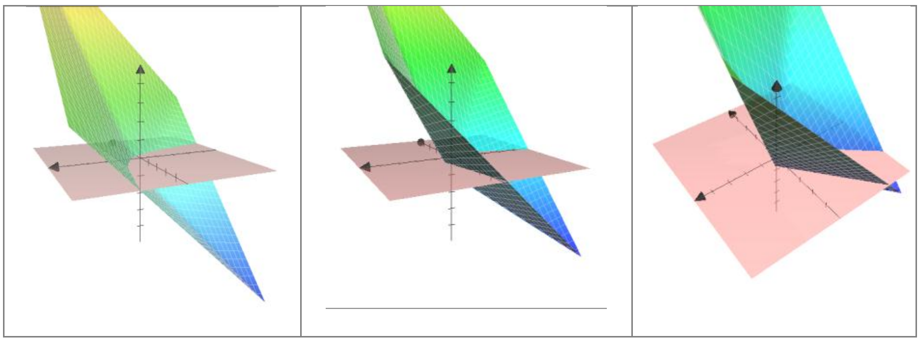
Şekil 3.4 Bir PKF'nin grafiği ve seviye kümesinin w_1 parametresine göre değişimi

Aynı fonksiyonda w_1 parametresi 0.1 olarak sabitlenerek w_2 parametresi için [-2,1] aralığındaki değişim Şekil 3.5 ile gösterilmektedir. Sol baştaki ve ortadaki şekillerde seviye kümesi üç hiperdüzlemin kesişimi ile oluşan açık bir dışbükey kümedir. Seviye kümesi, parametrenin 1'e yaklaşmasıyla daha dar bir alanı tanımlayan ve tamamen kapalı olan bir dışbükey kümeye dönüşmektedir.



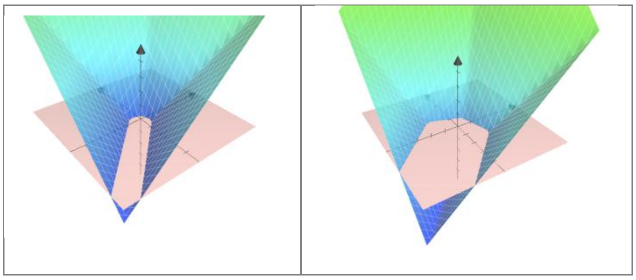
Şekil 3.5 Bir PKF'nin grafiği ve seviye kümesinin w_2 parametresine göre değişimi

Fonksiyondaki normlu terimin katsayısı ξ parametresinin değişimi [0,2] aralığında incelenmiş ve elde edilen grafikler Şekil 3.6 ile sunulmuştur. Bu parametrenin sıfır değerini alması, bir hiperdüzlem denklemi elde edilmesini sağlar. Parametrenin 2'ye doğru değişmesi ile de yine birden çok hiperdüzlemin kesişimi ile oluşan dışbükey kümelerin tanımladığı ayırıcı yüzeyler elde edilir.



Şekil 3.6 Bir PKF'nin grafiği ve seviye kümesinin ξ parametresine göre değişimi

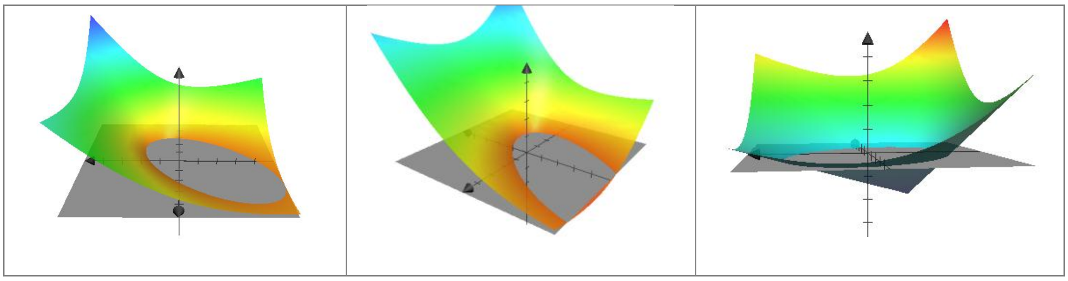
Fonksiyonun z ekseni üzerinde aşağı ve yukarı hareket etmesini sağlayan parametre γ parametresidir. Şekil 3.7 bu parametrenin 1 ve 2 değerleri için grafikler ve seviye kümeleri gösterilmektedir.



Şekil 3.7 Bir PKF'nin grafiği ve seviye kümesinin γ parametresine göre değişimi

***l_2* normu** : l_2 normu ile oluşturulan ve genel hali aşağıda verilen denklem (1.2.3) ile tanımlanan bir konik fonksiyonun grafiği Şekil 3.8 ile gösterilmiştir. l_2 normunun en önemli özelliği elipsoid şeklinde seviye kümelerinin oluşmasına imkan tanımasıdır.

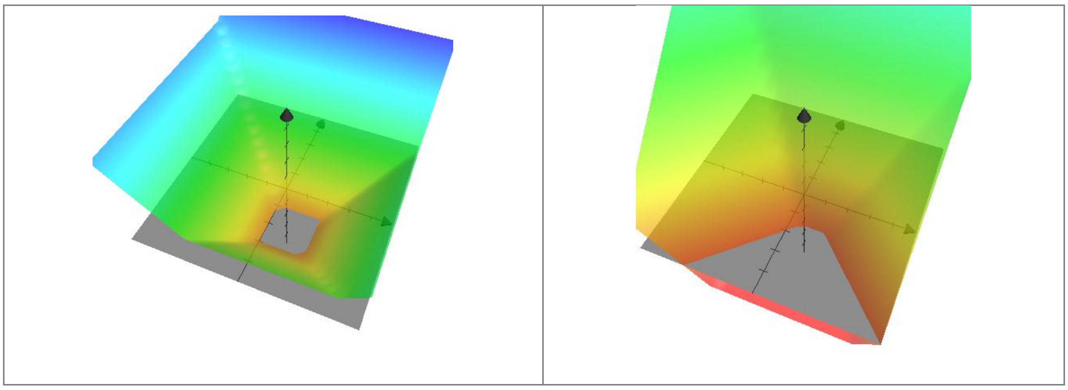
$$g_{l2}(x) = wx + \sigma\|x\|_2 - \gamma \qquad (1.2.3)$$



Şekil 3.8 l_2 normu ile oluşturulan bir konik fonksiyon

***l_{max}* normu** : $l_{\max}$ normu ile oluşturulan ve genel hali aşağıdaki denklem (1.2.4) ile tanımlanan bir konik fonksiyonun grafiği Şekil 3.9 ile gösterilmiştir. $l_{\max}$ normu, $l1$ normu ile elde edilmesi mümkün olmayan dışbükey kümelerin elde edimesini sağlamaktadır. Örneğin R^2'de l_1 normu ile elde edilmesi mümkün olmayan birim karelerin l_{max} normu ile elde edilmesi imkanı bulunmaktadır. l_{max} normu sınıflandırma problemlerinin çözümünde kullanılan hiper kutuların elde edilmesini sağlayan genel bir yapı sunmaktadır. Bu özelliği ile de sınıflandırma problemleri için dikkat çekicidir.

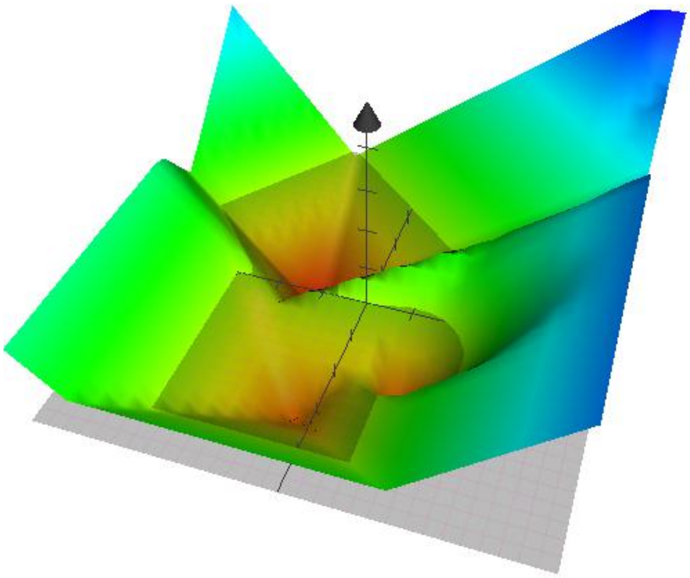
$$g_{l\max}(x) = wx + \phi\|x\|_{\max} - \gamma \qquad (1.2.4)$$



Şekil 3.9 l_{max} normu ile oluşturulan bir konik fonksiyon

l_1 , l_2 ve l_{max} normlarının birlikte kullanımı iki şekilde gerçekleştirilebilmektedir. Bunlardan biri, aşağıda denklem (1.2.5)'te tanımlandığı gibi yeni konik fonksiyonların tanımlanarak sınıflandırma amaçlı kullanılması, diğeri ise Şekil 3.10 ile de gösterildiği gibi ayırıcı fonksiyonu oluştururken farklı normlar ile elde edilmiş konik fonksiyonların noktasal minimumlarının kullanılmasıdır.

$$g(x) = wx + \xi\|x\|_1 + \sigma\|x\|_2 + \phi\|x\|_{\max} - \gamma \tag{1.2.5}$$



Şekil 3.10 Farklı normlar kullanılarak tanımlanan fonksiyonların noktasal minimumları ile oluşturulan bir fonksiyon.

3.3.4 Konik fonksiyonlarda ||.||2 normunun incelenmesi

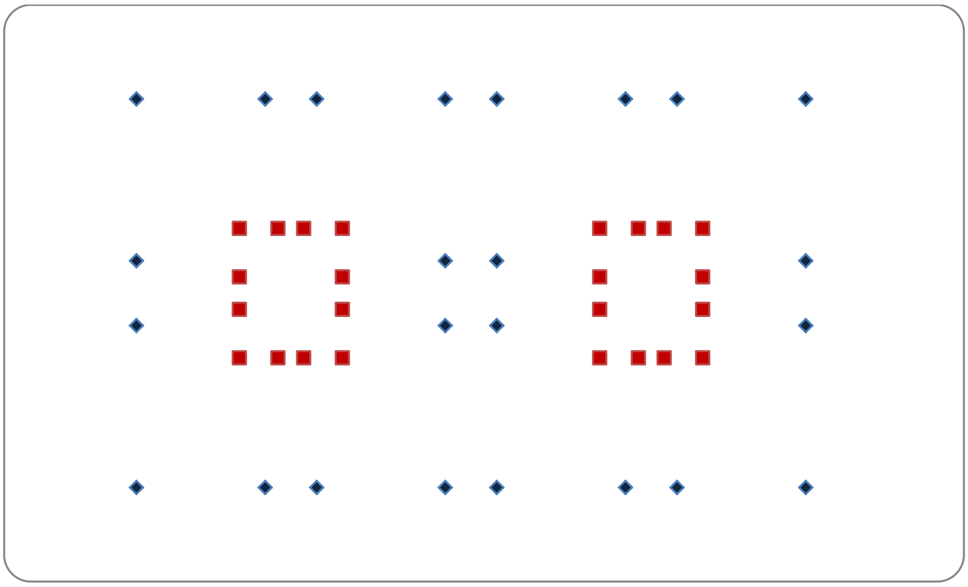
Polihedral konik fonksiyonlarda kullanılan l_1 normu yerine l_2 normunun kullanılması ile elde edilecek yeni fonksiyonların seviye kümelerinin sınıflandırma amacıyla kullanılması planlanmıştır. Bu kapsamda öncelikli olarak fonksiyonların genel yapısını grafikler yardımıyla incelemek için izleyen şekilde verilmiş olan örnek kullanılmıştır. Bu örnekte *A* ve *B* kümeleri

R^2 de tanımlanmış iki sonlu nokta kümesidir. Bu kümelerdeki veri noktaları Tablo 3.6 ile verilmiştir. *A* kümesine ait noktaların *B* kümesinin dışbükey örtüsü içinde yer aldığı Şekil 3.11'de gözükmektedir. Bu sebeple iki kümenin doğrusal ayırma ile tam olarak birbirinde ayrılması mümkün değildir.

Tablo 3.6 A ve B kümelerine ait noktaların koordinatları.

a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8	a_9	a_{10}	a_{11}	a_{12}	a_{13}	a_{14}	a_{15}	a_{16}	a_{17}	a_{18}	a_{19}	a_{20}	a_{21}	a_{22}	a_{23}	a_{24}
14.5	12	-0.5	-2	2	2	16	13.5	2	0.5	16	16	16	0.5	-0.5	-2	-2	2	12	-2	13.5	12	12	14.5
2	2	2	0.5	0.5	-2	-0.5	2	-0.5	2	0.5	2	-2	-2	-2	-0.5	2	2	-2	-2	-2	0.5	-0.5	-2

b_1	b_2	b_3	b_4	b_5	b_6	b_7	b_8	b_9	b_{10}	b_{11}	b_{12}	b_{13}	b_{14}	b_{15}	b_{16}	b_{17}	b_{18}	b_{19}	b_{20}	b_{21}	b_{22}	b_{23}	b_{24}
8	6	20	6	15	20	1	-1	-6	20	-6	-6	8	13	20	6	15	13	8	-1	-6	6	8	1
-6	-1	6	-6	6	1	6	6	1	-6	-1	-6	-1	6	-1	1	-6	-6	1	-6	6	6	6	-6

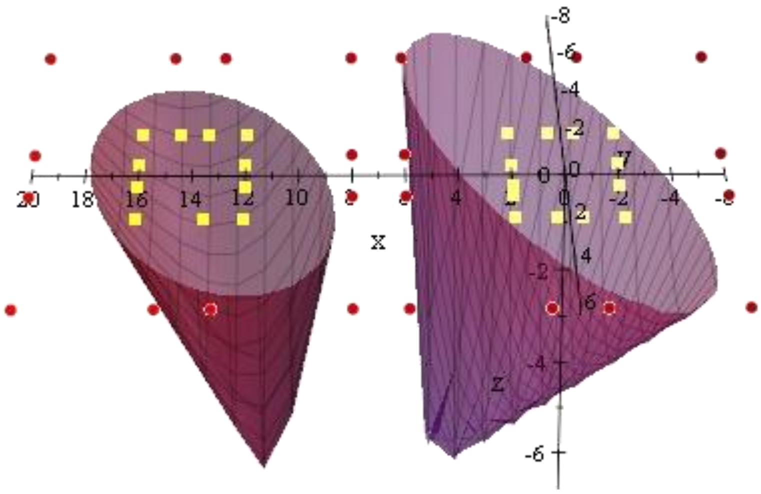


Şekil 3.11 *A* ve *B* kümelerine ait noktalar

A ve *B* kümelerini ayırmak için l_2 normu kullanılarak elde edilen ve genel yapısı $g(x) = wx + \xi\|x\|_2 - \gamma$ şeklinde olan fonksiyonlara ait parametreler Tablo 3.7 ile verilmiştir. Elde edilen bu iki fonksiyonun grafiği çizildiğinde Şekil 3.12'de görüldüğü gibi iki kümenin tam olarak birbirinden ayrılması sağlanabilmiştir.

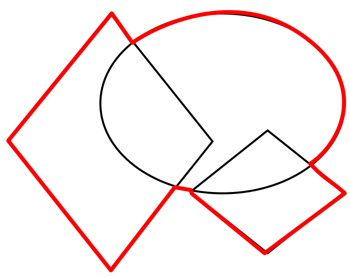
Tablo 3.7 l_2 normu ile elde edilen fonksiyonlar

	w_1	w_2	a_1	a_2	ξ	γ
$g_1(x)$	-7.03	0.88	10.92	3.4	1.7	5.03
$g_2(x)$	4.35	-5.81	4.93	-6.54	7.9	10.0



Şekil 3.12 l_2 normu ile oluşturulmuş ayırıcı fonksiyonlar.

Çoklu sınıflandırma problemlerinin çözümü için oluşturulacak olan konik fonksiyonlarda l_1 normu ile birlikte l_2 normunun da kullanılması, kümelerin birbirinden ayrılmasını sağlayacak dışbükey olmayan yarı düzgün (semi-smooth) yüzeyler oluşturulmasını sağlayacaktır. Böylece R^2'de Şekil 3.13 ile gösterilen ayırma yüzeylerine benzer dışbükey olmayan yarı düzgün yüzeylerin oluşturulması mümkün olacaktır.



Şekil 3.13 Dışbükey olmayan yarı-düzgün ayırma yüzeyi.

3.3.5 Test problemlerinin sonuçlarının analizi ve raporlanması

PKF algoritması ve İki Amaçlı Tamsayılı (İTS) Yaklaşım temelli ikili sınıflandırma yöntemleri kullanılarak **1e1** ve **1eh** yaklaşımları ile yeni çok sınıflı sınıflandırıcılar geliştirilmiştir. Yeni geliştirilen bu yöntemler 1e1-PKF, 1eh-PKF, 1e1-İTS ve 1eh-İTS olarak adlandırılmaktadır. Bu yöntemler Tablo 3.8 ile verilen ve literatürde yaygın olarak kullanılan çok sınıflı sınıflandırma problemleri ile test edilmiştir.

Tablo 3.8Çok sınıflı sınıflandırma problemleri

	Örnek Sayısı	Sınıf Sayısı	Özellik Sayısı
Iris	150	3	4
Wine	178	3	13
Glass	214	6	13
Vowel	528	11	10
Vehicle	846	4	18

Tablo 3.8'de ayrıca ele alınan test problemleri ile ilgili örnek, sınıf ve özellik sayıları da verilmektedir. Bu test problemleri ile ilgili kısa açıklamalar izleyen paragraflarda verilmektedir.

Iris Plant: Örüntü tanıma literatüründe bulunan veritabanlarından en çok bilineni Iris Plant veritabanıdır. Fisher'in çalışması bu alanda bir klasiktir ve günümüzdeki çalışmalarda hala referans verilmektedir. Veri kümesinde her birinde 50 örnek olan 3 sınıf bulunmaktadır. Her bir sınıf, bir iris bitkisi tipini işaret etmektedir. Bir sınıf, birbirinden doğrusal olarak ayrılamayan diğer iki sınıftan doğrusal olarak ayrılabilir.

***Wine*:** Bu veritabanında bulunan veriler, İtalya'nın aynı bölgesinde yetişen ancak üç farklı biçimde ekilen üzümlerden elde edilen şarapların kimyasal bir analizinin sonuçlarıdır.

Glass : Cam veri kümesi, bir cam örnekleminin öz yapısını belirlemek için kullanılmıştır. Cam türlerinin sınıflandırıldığı bu çalışma kriminolojik araştırma ile güdülenmiştir.

Vovel: Sesli harfler veri kümesindeki veriler California Üniversite'sinden elde edilmiştir.

PKF algoritması temelindeki çoklu sınıflandırma yaklaşımları için elde edilen sonuçlar Tablo 3.9 ile iki amaçlı tamsayılı programlama yaklaşımı kullanılarak elde edilen sonuçlar ise Tablo 3.10 ile verilmiştir. PKF algoritması eğitim aşamasında herhangi iki kümeyi tam olarak ayırabildiği için, Tablo 3.9'de verilen 1e1-PKF ve 1eh-PKF yöntemlerinin tamamında eğitim başarısı %100 olmuştur. İki amaçlı tamsayılı yaklaşım temelindeki çoklu sınıflandırma yöntemlerinde ise eğitim ve test başarıları Tablo 1.5.3'de birlikte verilmiştir.

En sık karşılaşılan çok sınıflı sınıflandırma problemlerinden biri olan Iris için literatüde %81 ile %98.67 arasında değişen, çok farklı sonuçlar yer almaktadır.. Iris veri kümesi için bu çalışmada elde edilen en iyi on-kez çapraz doğrulama başarımı %97.33, 1eh-ITS yaklaşımı ile elde edilmiştir.

Wine veri kümesi de Iris gibi sık kullanılan veri kümelerindendir. Literatürde bu veri kümesi üzerinde elde edilmiş test başarımları %88.44 ile %100 arasında değişmektedir. 1eh-ITS yaklaşımı ile wine veri kümesi için %96.63 test başarımı elde edilmiştir.

Glass veri kümesi için doğrusal yüzeyler kullanan yaklaşımlar için test başarımları %54.67 ile %67.2 arasında gerçekleşmiştir. En yüksek başarı doğrusal olmayan kernel fonksiyonları kullanarak %73.832 olarak raporlanmıştır. Glass veri kümesi için en iyi ikinci çapraz doğrulama başarımı %72.43, 1eh-İTS yaklaşımı ile elde edilmiştir.

Vowel veri kümesi için 1eh-İTS yaklaşımı ile elde edilen sonuç %92.05 olmuştur. Bu sonuç doğrusal yüzeyler veya kernel fonksiyonu kullanan yaklaşımlar ile elde edilen sonuçlardan daha iyidir. Ancak doğrusal olmayan kernel fonksiyonlarının kullanıldığı yaklaşımlar ile daha iyi sonuçlar elde edilmiştir.

Tablo 3.9 PKF algoritması temelli yöntemler ile on kez çapraz doğrulama test sonuçları

	1e1-PKF	**1eh-PKF**
Iris	93.33	96.67
Wine	92.7	96.63
Glass	67.29	65.42
Vowel	87.31	87.12
Vehicle	79.91	78.28

Tablo 3.10 İki amaçlı tamsayılı PKF yaklaşımı için on kez çapraz doğrulama eğitim ve test sonuçları

	1e1-İTS		**1eh-İTS**	
	Eğitim	**Test**	**Eğitim**	**Test**
Iris	100	94.67	98.59	97.33
Wine	100	92.7	100	96.63
Glass	95.07	72.43	100	70.6
Vowel	99.96	87.12	99.87	92.05
Vehicle	94.68	82.03	94.85	79.79

Çok sınıflı problemlerin çözümü için PKF temelli yeni yaklaşımlar önerilmiştir. Tüm veri kümeleri için PKF algoritması ile 2, iki amaçlı tamsayılı PKF yaklaşımı ile 2 olmak üzere toplam dört model ile deneyler yapılmıştır ve elde edilen sonuçlar sunulmuştur. PKF algoritması temelli yöntemler ile elde edilen sonuçların tümünde eğitim başarımı 100 olarak gerçekleşmektedir. İki amaçlı tamsayılı yaklaşım ile elde edilen sonuçlar, PKF algoritmasına göre genellikle daha iyi çıkmıştır. Ayrıca eğitim ve test başarımları arasındaki fark azalmış, böylece genelleştirme başarımı artmıştır. Literatürde en iyi sonuçlar Hsu ve Lin (2002) tarafından, doğrusal olmayan kernel fonksiyonları kullanan destek vektör makineleri ile elde edilmiştir. Hsu ve Lin, her test problemi için en iyi kernel fonksiyonunu belirlemek üzere, fonksiyon parametrelerinin farklı değerlerine göre 225 model çözdürmüş ve elde ettikleri en iyi sonuçları yayınlamışlardır. PKF temelli yaklaşımların hiçbirinde DVM'lerin kernel düşüncesi kullanılmamıştır. Buna rağmen doğrusal olamayan DVM'ler ile rekabet edebilecek sonuçlar elde edilmiştir.

Projede l1, l2 ve lmax normlarının farklı kombinasyonlarının lineer fonksiyona eklenmesi ile altı farklı fonksiyon tanımlanmıştır. PKF algoritması, yeni tanımlanan bu fonksiyonlar ile çalışacak şekilde uyarlanmıştır. PKF algoritmasının yeni hali, literatürde iyi bilinen ve daha önce l1 normu için denemeleri yapılan çok sınıflı test problemleri üzerinde denenmiştir. Test veri kümeleri ile ilgili bilgiler Tablo 3.11 ile verilmiştir. PKF algoritması temelli bire karşı bir ve bire karşı hepsi yaklaşımları ile elde edilen sonuçlar sırasıyla Tablo 3.12 ve Tablo 3.13 ile özetlenmiştir.

Tablo 3.11 Ele alınan çok sınıflı problemler için örnek, özellik ve sınıf sayıları

	01 Iris	02 Wine	03 Glass	04 Vowel	05 Vehicle
Örnek sayısı	150	178	214	528	846
Sınıf sayısı	3	3	6	11	4
Özellik sayısı	4	13	13	10	18

Tablo 3.12 Bire karşı bir (1e1) yaklaşımı ile elde edilen test veri kümelerine ait sonuçlar

	Iris	Wine	Glass	Vowel	Vehicle
01-1e1-PKF-l_1	93.33	92.7	67.29	87.31	79.91
l_2	93.33	94.38	62.62	85.42	81.56
(l_1 , l_2)	94.00	96.07	67.76	85.61	81.80
(l_1, l_{max})	96.67	94.94	65.89	85.04	80.14
(l_1, l_2 , l_{max})	94.00	96.07	64.95	85.23	80.26
l_{max}	96.67	94.94	64.49	85.23	81.09

Test veri kümeleri üzerinde yapılan denemeler sonucunda yeni konik fonksiyonlar içinde tamamen diğerlerine üstün gelen bir fonksiyona rastlanmamıştır. Fakat yukarıdaki tablodan da görüldüğü gibi farklı fonksiyonlar farklı veri kümeleri için daha başarılı olmuştur. Bazı fonksiyonlar için oldukça dikkat çekici sonuçlar elde edilmiştir. "İris" veri kümesi için 1e1-PKF yaklaşımı ile l1 normlu fonksiyonlar kullanılarak elde edilen başarı oranı 93.33 olmuş, fakat bu veri kümesi için (l1,lmax) ve lmax normlu fonksiyonlar kullanılarak 96.67 gibi yüksek bir başarı oranı elde edilmiştir. Benzer şekilde "Wine" veri kümesi için 96.07 başarı oranı (l1,l2) ve (l1, l2, lmax) normlu fonksiyonlar ile, "Vehicle" veri kümesi için 81.80 başarı oranı (l1,l2) normlu fonksiyon ile elde edilmiştir.

Tablo 3.13 Bire karşı hepsi(1eH) yaklaşımı ile elde edilen test veri kümelerine ait sonuçlar

	Iris	Wine	Glass	Vowel	Vehicle
01-1eH-PKF-l_1	96.67	96.63	65.42	87.12	78.28
l_2	95.33	95.51	64.95	84.09	76.12
(l_1 , l_2)	94.67	95.51	68.69	84.47	77.66
(l_1, l_{max})	96.67	93.26	71.5	82.77	77.66
(l_1, l_2 , l_{max})	94.00	93.26	66.36	83.90	78.37
l_{max}	95.33	93.26	64.02	83.90	76.83

Tablo 3.13'da görüldüğü gibi, 1-e-h yaklaşımı ile "Iris", "Wine", "Vowel" veri kümeleri için en yüksek başarı oranları l_1 normlu PKF algoritması ile elde edilmiştir. "Glass" veri kümesi için en yüksek başarı oranı l_1 *ve* l_{max} normlarının kombinasyonu, "Vehicle" veri kümesi için de en yüksek başarı oranı l_1, l_2 ve l_{max} normlarının kombinasyonu ile elde edilmiştir.

Bu sonuçlar, projede geliştirilen yeni sınıflandırma algoritmalarının önemini göstermektedir.

4 İMKB'DE İŞLEM GÖREN ŞİRKETLERİN DERECELENDİRİLMESİ İÇİN GELİŞTİRİLEN BİR KARAR DESTEK SİSTEMİ

4.1 Veri Tabanının Oluşturulması

Şirketlerin değerlendirilmesinde kullanılacak finansal oranların hesabında İMKB'deki şirketlerin bir alt kümesi olan Ulusal Sınaî Endeksi'nde yer alan şirketler kullanılmış ve Finansal oranların hesaplanmasında zaman aralığı, 2002–2007 yılları olarak belirlenmiştir. 2007 yılında bu endekste 151 şirket olmasına rağmen, 2002–2007 yılları arasında verileri tam olan şirket sayısı 132 adet olarak belirlenmiş ve çalışmada kullanılmak üzere bu şirketler seçilmiştir. Bu 132 şirketin her biri için 12 aylık bilânçolarında verilen bilgiler yardımıyla 19 adet finansal oran hesaplanmış ve Ek Tablo 1- Ek Tablo 6 ile verilmiştir.

Ayrıca 132 adet şirketin, hisse senedi getirilerine ve karlılıklarına göre sınıflandırılmasına ihtiyaç duyulmuştur. Daha önce belirlenen 132 adet şirketin hisse senedi getirilerine göre değerlendirilmesi gerekmektedir. Bu amaçla IMKB'de işlem gören 132 adet şirketin 2002-2007 yılları arasındaki yüzdelik aylık getiri oranları, IMKB'nin resmi internet sitesinden (http://www.imkb.gov.tr/sirket/fiyat_getiri.htm) sağlanıp düzenlenmiş ve bu veriler kullanılarak elde edilen yıllık getiriler Ek Tablo 7 ile verilmiştir. Getirilerin hesabında IMKB'nin kullandığı formülasyon, kullanılmıştır.(http://www.imkb.gov.tr/sirket/fiyat_getiri_aciklama.htm).

Bahsi geçen 132 adet şirket için 2002-2007 yılları arasında hesaplanan 19 finansal oran ve getirileri oranlarının saklanacağı bir veri tabanı oluşturulmuştur.

4.1.1 IMKB'de işlem gören hisse senedlerinin aylık getirilerinin hesabı

IMKB'de işlem gören hisse senedlerinin aylık getirileri, bir hisse senedinin bir ay boyunca elde tutulması sonucunda elde edilen getiriyi göstermekte olup aşağıdaki formüle göre hesaplanmıştır.

Parametreler:

g_i : *i.* aya ait getiri oranı,
f_i : *i.* aya ait en son kapanış fiyatı,
f_{i-1} : *i.* aydan bir önceki aya ait en son kapanış fiyatı,
BDL : Ay içinde alınan bedelli hisse adedi,
BDZ : Ay içinde alınan bedelsiz hisse adedi,
r : Rüçhan hakkı kullanma fiyatı,
t : Ay içinde 1.000,-TL/1 TL nominal değerli bir hisse senedine ödenen net temettü tutarı olmak üzere,

$$g_i = \frac{f_i \times (BDL + BDZ + 1) - r \times BDL + t - f_{i-1}}{f_{i-1}}$$

denklemi ile hesaplanmıştır. Yukarıdaki formülasyona göre hesaplanan getiri oranlarından hareketle şirketlere ait hisse senetlerinin yüzdelik yıllık bileşik getiri ve yıllık getiri standart sapmaları hesaplanmıştır.

4.1.2 Şirketlere ait hisse senedlerinin yıllık bileşik getiri ve yıllara göre aylık getiri standart sapmalarının hesabı

Yıllık bileşik getiri, bir hisse senedinin yıl boyunca her ay sonunda satılıp tekrar alınması sonucunda başlangıç dönemindeki değerinin kaç katına ulaştığını göstermekte olup aylık getiri verileri kullanılarak aşağıdaki formüle göre hesaplanmıştır.

Parametreler:

n : Dönem (ay) sayısı,
G_n : *n*. ay sonuna kadar bileşik getiri,
g_i : *i.* aya ait getiri olmak üzere,

$$G_n = \prod_{i=1}^{n} (1 + g_i)\,.$$

şeklinde hesaplanmıştır. Standart Sapmalar ise aylık yüzdelik getiri oranı ortalamaları,

$$\bar{G}_n = \frac{\sum_{i=1}^{n} g_i}{n}$$

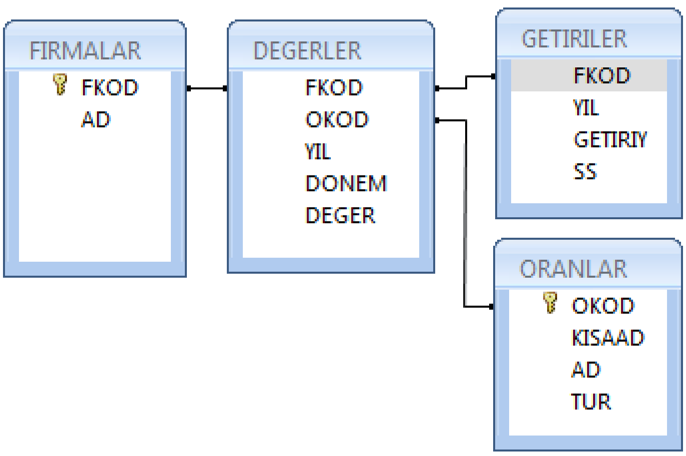
olmak üzere

$$SS_n = \frac{\sum_{i=1}^{n} (g_i - \bar{G}_n)^2}{n-1}$$

denklemi ile hesaplanmıştır. 132 şirketin 2002–2007 yıllarına ait *n*=12 için hesaplanan yıllık bileşik getiri ve yıllık getiri standart sapması verileri, Ek Tablo 7 ile verilmiştir.

4.1.3 Finansal oranlar ve getiri oranları veritabanı

İMKB ulusal sınaî indeksinde yer alan 132 adet şirketin 2002-2007 yılları arasındaki bilançolarından elde edilen 19'ar Finansal oranın ve yine aynı dönemlerdeki getiri oranlarının sistematik bir şekilde depolanması amacı ile oluşturulan veritabanında bulunan tablolar ve tabloların arasındaki ilişkiler Şekil 4.1 ile gösterilmektedir.



Şekil 4.1 Finansal oranlar ve getiri oranları veritabanı tabloları ve ilişkiler

Bu veri tabanındaki ORANLAR, GETIRILER, DEGERLER ve FIRMALAR tabloları ile ilgili örnek ekran çıktıları sırasıyla Şekil 4.2, Şekil 4.3, Şekil 4.4 ve Şekil 4.5 ile gösterilmektedir.

OKOD	KISAAD	AD	TUR
K01	SÖSO	Sermaye – Özsermaye Oranı	Kaldıraç
K02	UBDSO	Uzun Vadeli Borçlar – Devamlı Sermaye Oranı	Kaldıraç
K03	UBAO	Uzun Vadeli Borçlar – Aktifler Oranı	Kaldıraç
K04	KBTBO	Kısa Vadeli Borçlar – Toplam Borçlar Oranı	Kaldıraç
K05	ÖÇ	Özsermaye Çarpanı	Kaldıraç
K06	BÖO	Borç-Özsermaye Oranı	Kaldıraç
K07	BAO	Borç-Aktifler Oranı	Kaldıraç
L01	CO	Cari Oran	Likidite
L02	LO	Likitide Oranı	Likidite
L03	NO	Nakit Oranı	Likidite
L04	NSAO	Net İşletme Sermayesi – Aktifler Oranı	Likidite
L05	LAAO	Likit Aktifler – Aktifler Oranı	Likidite
M01	MDVÖO	Maddi Duran Varlıklar – Özsermaye Oranı	Mali Yapı
M02	MDVUBO	Maddi Duran Varlıklar – Uzun Vadeli Borçlar Oranı	Mali Yapı
M03	MDVDSO	Maddi Duran Varlıklar – Devamlı Sermaye Oranı	Mali Yapı
M04	DVUBO	Duran Varlıklar – Uzun Vadeli Borçlar Oranı	Mali Yapı

Şekil 4.2 Oranlar Tablosu

ORANLAR tablosu, oran kodunu gösteren OKOD, oranın kısa adını gösteren KISAAD, oranın adını gösteren AD, ve oranın türü gösteren TUR alanlarından oluşmaktadır. Bu tablo her orana özel verilerin bulunduğu statik bir tablodur. Proje kapsamında ihtiyaçlar dâhilinde gerekli görülürse yeni alanlar eklenerek tablo geliştirilebilir. Örneğin oranların farklı bir sınıflaması yapılmak istenirse yeni bir alan eklenerek bu ihtiyaç karşılanabilir. Bazı oranların bazı sektörlere göre daha farklı anlamları olması durumunda sektörel bazda benzer bir ihtiyaç karşımıza çıkabilir.

FKOD	YIL	GETIRIY	SS
ADANA	2007	-1.5896	8.5624
ADEL	2007	-7.4184	58441194
SASA	2007	-16.8675	4.5901
AFYON	2007	-0.8400	10.0761
AKALT	2007	20.2893	16.2420
AKCNS	2007	-10.8361	9.1487
AEFES	2007	28.3633	8.8690
AKSA	2007	-22.8192	9.5217
AKIPD	2007	19.2308	9.8845
ALCAR	2007	-14.9968	6.6449
ALKA	2007	38.3838	9.5158
ALTIN	2007	128.6585	16.2663
ANACM	2007	6.0745	8.2379
ARSAN	2007	18.1818	11.4889
ASUZU	2007	9.5148	6.9603
ARCLK	2007	1.2539	9.0093
ATEKS	2007	-0.8197	4.1318
AYGAZ	2007	60.4738	8.8886
BAGFS	2007	162.7381	10.2342
BAKAB	2007	-28.6486	6.2541

Şekil 4.3 Getiriler Tablosu

GETIRILER tablosu, firma kodunu gösteren FKOD, getirilerin hangi yıla ait olduğunu gösteren YIL, getiri yüzdelerini gösteren GETIRIY ve bölüm 4.1.2'de anlatıldığı gibi hesaplanan standart sapmayı gösteren SS alanlarından oluşmaktadır. Bu tabloda firmaların yıllar bazında getiri yüzdeleri ve standart sapmaları yer alır. Bu tabloda mevcut durumda 2002-2007 yılları arasında, 6 yıl, 132 firma için toplam 792 kayıt bulunmaktadır.

FKOD	OKOD	YIL	DONEM	DEGER
ADANA	L01	2002	12	2.21692827277904
ADANA	L02	2002	12	1.44625606702052
ADANA	L03	2002	12	0.910183202771706
ADANA	L04	2002	12	0.226028014180279
ADANA	L05	2002	12	26.8622559058733
ADANA	M01	2002	12	40.3473881264006
ADANA	M02	2002	12	656.738180350672
ADANA	M03	2002	12	38.0120769361561
ADANA	M04	2002	12	1248.12164703512
ADANA	K01	2002	12	31.6443495200035
ADANA	K02	2002	12	5.78801081975456
ADANA	K03	2002	12	4.71296587182554
ADANA	K04	2002	12	79.7610540279939
ADANA	K05	2002	12	1.30355351748588
ADANA	K06	2002	12	30.3553517485877
ADANA	K07	2002	12	23.2866171901658
ADEL	L01	2002	12	2.23985666448328
ADEL	L02	2002	12	0.495515444706334
ADEL	L03	2002	12	0.11960789622237

Şekil 4.4 Değerler Tablosu

DEGERLER tablosu, firma kodunu gösteren FKOD, oran kodunu gösteren OKOD, oranların hangi yıla ve döneme ait olduğunu gösteren YIL ve DONEM ve finansal oranın değerini gösteren DEGER alanlarından oluşmaktadır. Bu tablo firmaların yıllar ve dönemler bazında 19 farklı finansal oranına ait değerleri içermektedir. Bu tabloda mevcut durumda 2002-2007 yılları arasında, 6 yıl, 132 firma için, 19 oran olmak üzere toplam 15048 kayıt bulunmaktadır.

FKOD	AD
ADANA	ADANA ÇİMENTO SANAYİİ T.A.Ş.
ADEL	ADEL KALEMCİLİK TİCARET VE SANAYİ A.Ş.
AEFES	ANADOLU EFES BİRACILIK VE MALT SANAYİİ A.Ş.
AFYON	AFYON ÇİMENTO SANAYİ T.A.Ş.
AKALT	AK-AL TEKSTİL SANAYİİ A.Ş.
AKCNS	AKÇANSA ÇİMENTO SANAYİ VE TİCARET A.Ş.
AKIPD	AKSU İPLİK DOKUMA VE BOYA APRE FABRİKALARI T.A.Ş.
AKSA	AKSA AKRİLİK KİMYA SANAYİ A.Ş.
ALCAR	ALARKO CARRIER SANAYİ VE TİCARET A.Ş.
ALKA	ALKİM KAĞIT SANAYİ VE TİCARET A.Ş.
ALTIN	ALTINYILDIZ MENSUCAT VE KONFEKSİYON FAB.A.Ş.
ANACM	ANADOLU CAM SANAYİİ A.Ş.
ARCLK	ARÇELİK A.Ş.
ARSAN	ARSAN TEKSTİL TİCARET VE SANAYİ A.Ş.
ASUZU	ANADOLU ISUZU OTOMOTİV SANAYİ VE TİCARET A.Ş.
ATEKS	AKIN TEKSTİL A.Ş.

Şekil 4.5 Firmalar Tablosu

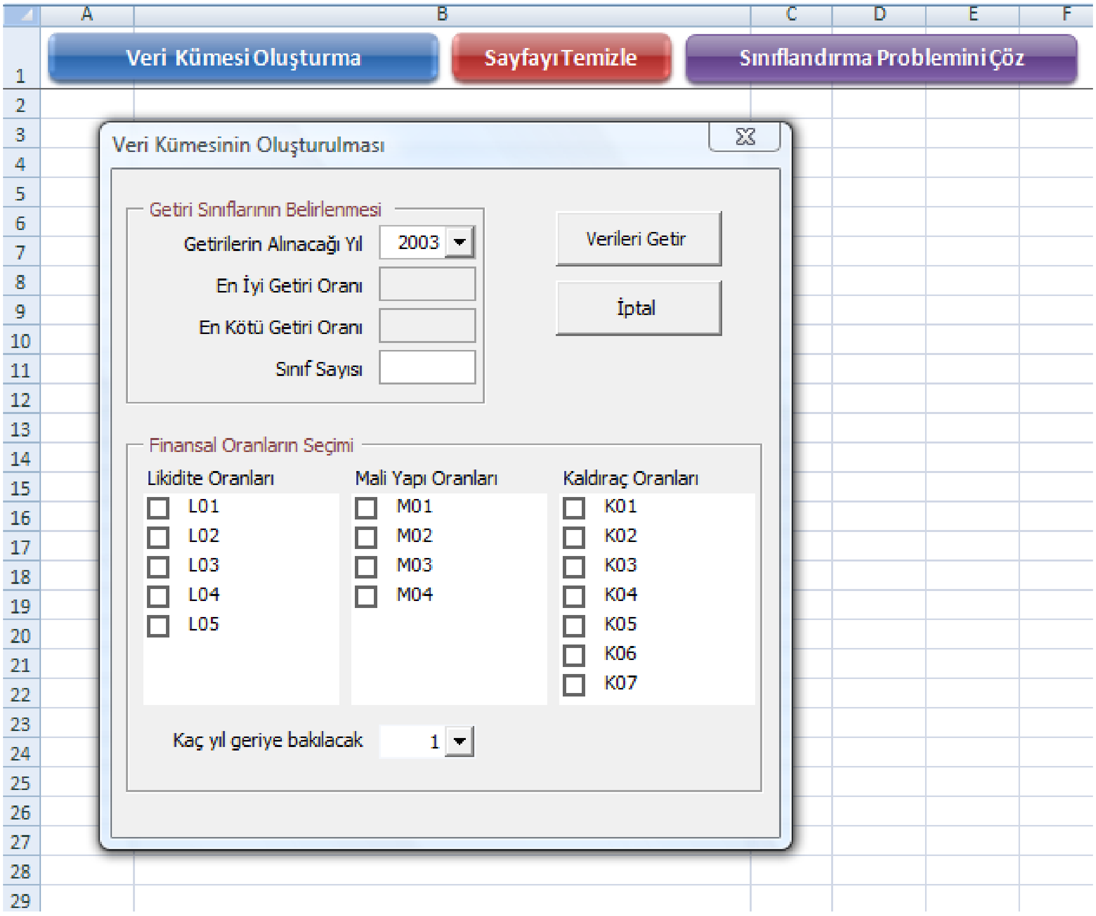
FIRMALAR tablosu, firma kodunu gösteren FKOD ve firmanın adını gösteren AD alanlarından oluşmaktadır. Bu tablo da ORANLAR tablosu gibi statik yapıdadır. İhtiyaçlar dahilinde gerekli görülmesi halinde tabloya yeni alanlar eklenerek genişletilmesi mümkün olabilir.

Veritabanı sayesinde bilgilere ulaşım daha hatasız ve hızlı bir şekilde olabilecektir. Ayrıca burada bulunan bilgiler farklı uygulamalara da hizmet edecek şekilde oluşturulmuştur. Finansal oranlar üzerine yapılacak herhangi bir çalışmada bu bilgilerin tekrar hazırlanması maliyeti ve külfetine katlanmak zorunda kalınmayacaktır.

4.2 Uygulama problemlerinin hazırlanması ve algoritmalarda denenmesi

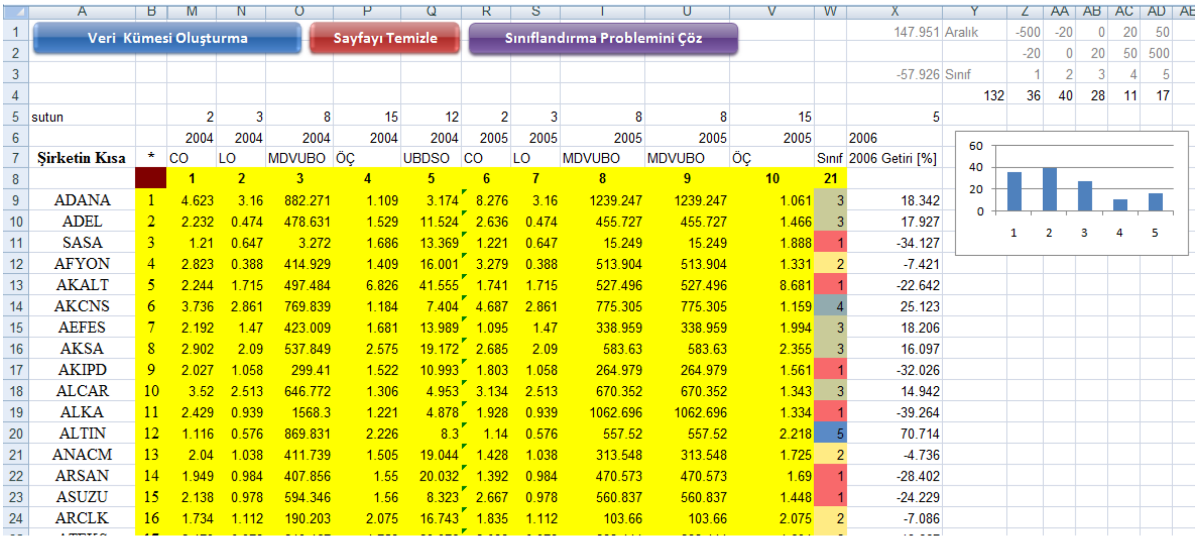
Sınıflandırma problemlerinin çözümünde kullanılacak olan algoritmalar GAMS aracılığı ile kodlanmıştır. Oluşturulan GAMS dosyalarının farklı problemleri çözebilmesi için probleme ait parametreler metin dosyaları aracılığı ile model dosyasına eklenir. GAMS modelinin okuduğu formatta veri dosyaları oluşturmak üzere Excel ortamında yeni bir sistem tasarlanmıştır. Bu sistem sayesinde kullanıcı kendi tercihleri doğrultusunda bir sınıflandırma problemini elde edebilmektedir. Oluşturulan bir kullanıcı formu ile hangi döneme ait getirilerin alınacağı, hangi finansal oranların kullanılacağı ve geriye dönük kaç dönem için verilerin alınacağı belirlenir. Kullanıcının bu form aracılığıyla belirttiği tercihlere göre Access veritabanına bağlanılarak gerekli veriler Excel ortamına çekilir. Excel sayfasındaki bu veriler GAMS modeli için uygun formatta otomatik olarak düzenlenir. Kullanıcı bu verileri metin dosyası formatında GAMS modeline göre kaydedebilir. Bu sayede, seçilen parametrelere göre özel oluşturulmuş problem GAMS'te çözülmek üzere hazırlanmış olur.

Şekil 4.6, sınıflandırma probleminin GAMS'te çözümüne uygun veri kümesi oluşturmak için tasarlanan Excel dosyasını ve "Veri Kümesi Oluşturma" kullanıcı formunu göstermektedir. Bu form sayesine önceki paragrafta bahsedilen tüm kullanıcı tercihleri alınır. Daha sonra "Verileri Getir" butonuna basıldığında tercihlere göre ilgilenilen veriler bu Excel sayfasına getirilir. Henüz sayfadaki ilk satırda bulunan üçüncü düğme, "sınıflandırma problemini çöz" düğmesi, problemin otomatik çözümünü yapmamaktadır. Bu düğmeye basıldığında, sadece sayfada bulunan veriler GAMS modelinde kullanılacak formatta bir metin dosyasına aktarılmaktadır. Önümüzdeki altı aylık dönemde, bu düğmenin seçilen problemi otomatik olarak GAMS'i çağırarak çözecek ve sonuçlarını Excel'e alacak şekilde düzenlenmesi planlanmaktadır. Bu sayede geliştirilen sistem veritabanı, model tabanı, kullanıcı arayüzleri ve kullanıcıları ile tam bir karar destek sistemi olacaktır.



Şekil 4.6 "Veri Kümesi Oluşturma" kullanıcı formu.

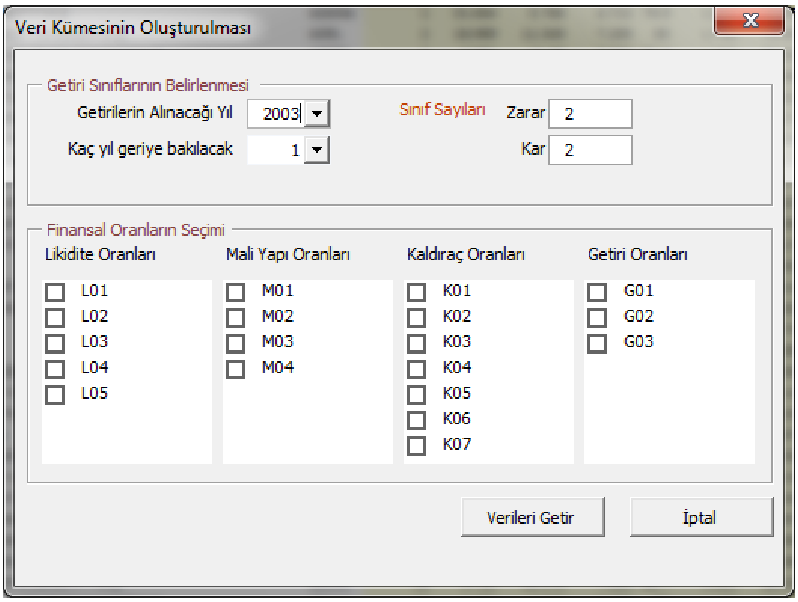
Kullanıcının belirlediği tercihlere göre oluşturulmuş bir örnek veri kümesi Şekil 4.7 ile gösterilmektedir. Verilerin Excel ortamına alınmasının ardından getirilere göre sınıflar belirlenerek veri kümesi oluşturulur. Sınıf sayısı ve sınıf aralıkları girildiğinde getiri yüzdelerinin bulunduğu sütuna göre getiri sınıfları otomatik olarak oluşturulmaktadır. Sınıf aralıklarının belirlenmesinin ardında veriler GAMS modeline uygun halde kaydedilir ve GAMS modeli çözdürülür.



Şekil 4.7 Örnek veri kümesi

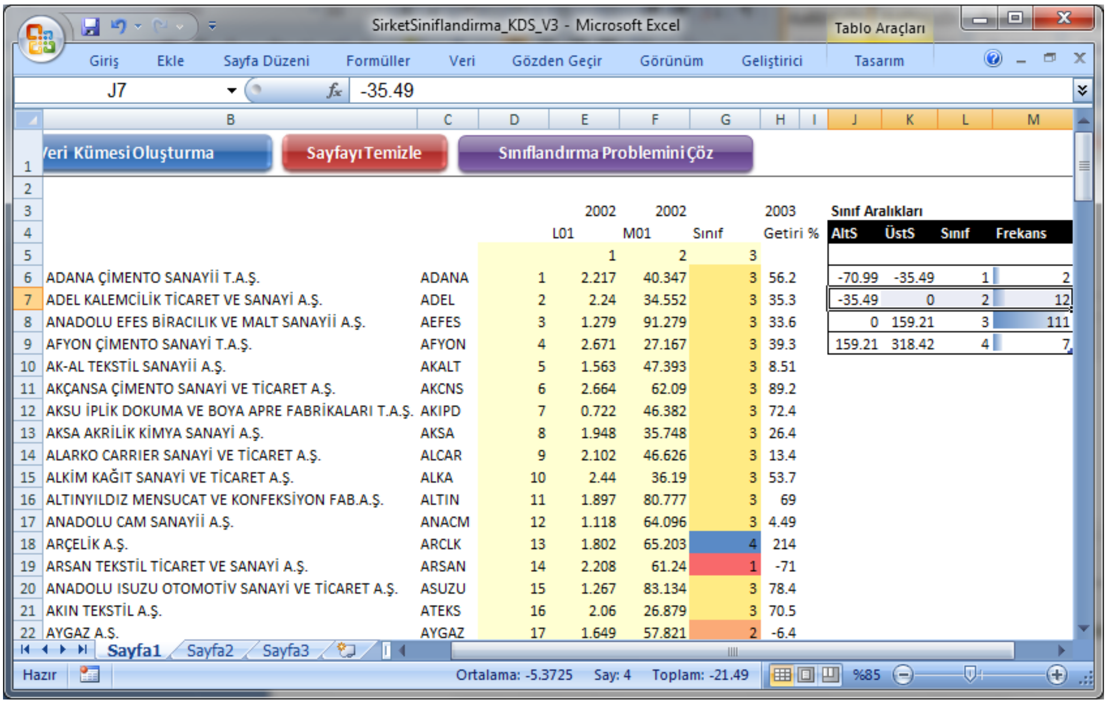
Uygulama problemlerinin tasarlanan matematiksel modellerde denemelerini sistematik olarak gerçekleştirebilmek için tasarlanan VBA temelli Excel dosyası üzerinde bazı geliştirmeler yapılmıştır. Bu geliştirmeler kısaca izleyen paragraflarda açıklanmaktadır.

Veri kümesinin oluşturulması Şekil 4.8 ile verilen arayüz sayesinde gerçekleştirilmektedir. Bu arayüzde kullanıcıya yeni finansal oranları seçebilme imkânı tanınmaktadır. Bunun yanı sıra, veri kümesinin çok önemli bir kısmını oluşturan getiri sınıflarının belirlenmesinde sınıf sayısının girilmesi için iki yeni metin kutusu arayüze eklenmiştir. Bu metin kutularından bir tanesi getiri yüzdelerine göre zarar eden şirketlerin kaç sınıfta ele alınacağını belirlemektedir. Diğer metin kutusu ise kar eden şirketler için sınıf sayısının alınmasında kullanılmaktadır.



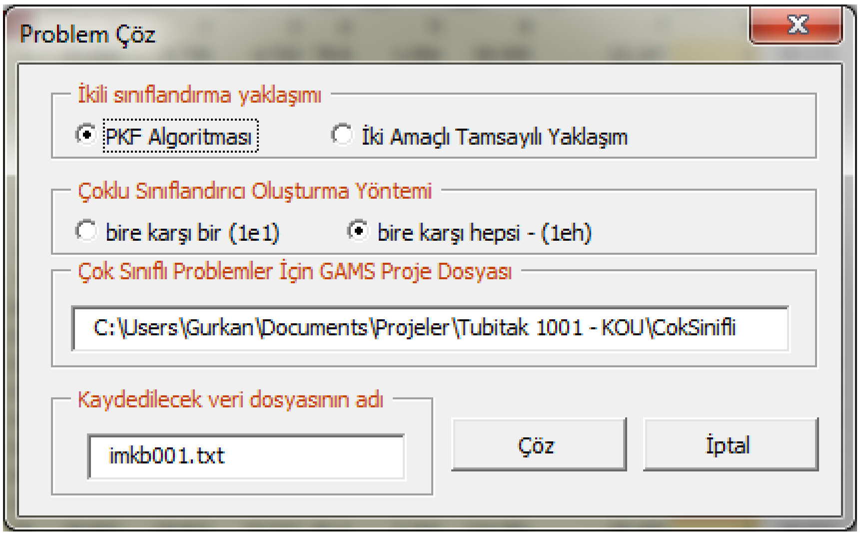
Şekil 4.8 Veri kümesi oluşturma arayüzü

Veri kümesi oluşturma arayüzünde gerçekleştirilen değişiklerin yanı sıra, verilerin Access veritabanından alınması ve Excel sayfasına getirilerek veri kümesinin oluşturulmasında da bazı değişiklikler gerçekleştirilmiştir. Verileri hatasız bir şekilde GAMS'te geliştirilen modellere aktarabilmek için Excel sayfasına, Şekil 4.9 ile gösterildiği gibi, satır ve sütun numaraları, hangi şirketin hangi sınıfa ait olduğunu gösteren "Sınıf" sütunu ve sınıf aralıkları eklenmiştir. Oluşturulan bu Excel sayfasının dinamik bir şekilde kullanılabilmesi için bahsedilen eklemeler yapılırken doğrudan değerleri yazdırmak yerine, getiri yüzdesine göre sınıfları belirleyen ve hangi sınıfta kaç şirket olduğunu sayan formüllerden yararlanılmıştır. Bu sayede kullanıcı isterse sınıf ararlıklarının alt sınırlarını değiştirerek yeni uygulama problemleri elde edebilmektedir.



Şekil 4.9 Bir uygulama problemi örneği

Excel sayfasında yapılan geliştirmelerden sonuncusu ise elde edilen uygulama probleminin GAMS programında doğrudan çözülmesini sağlayan bölümdür. Şekil 4.10 ile gösterilen "Problem Çöz" formu; çözüm yöntemi ile ilgili bilgilerin, GAMS proje dosyasının bulunduğu yerin ve uygulama problemi için GAMS'in kullanacağı "include" dosyanın adının kullanıcıdan alınmasını sağlar. "Çöz" düğmesine basıldığında, Excel'de bulunan veriler ve probleme ait bilgiler GAMS modellerinin okuyacağı formatta bir metin dosyasına yazdırılır ve ilgili algoritma çalıştırılır. Sonuçlar hem gams veri değiştirme dosyasına (gdx) hem de bir metin dosyasına yazdırılır.



Şekil 4.10 “Problem Çöz” arayüzü

Yapılan geliştirmeler sayesinde veritabanı, modelleri ve kullanıcı arayüzleri ile getirilere göre şirketlerin sınıflandırması için bir karar destek sistemi ortaya çıkarılmıştır.

IMKB’de İşlem Gören 132 Şirket’in Yıllık Hisse Senedi Getirilerine göre Çoklu Sınıflandırılmasında PKF algoritması temelli yaklaşımların, denenmesi için test problemlerinin oluşturulması gerekmektedir. Finansal oranlar, dört grupta ele alınmıştır. Bunlar, Likitide Oranları (5 adet), Mali Yapı Oranları (4 adet), Kaldıraç Oranları (7 adet) ve Piyasa Değerini Ölçen Oranlardır (3 adet). Test problemlerini tanımlamada kullanılan parametreler aşağıda verilmiştir.

t: Hisse senedi getirilerinin hesaplandığı yıl (*t* = 2002,…, 2007).

m: Hisse senedi getirilerinin ayrıldığı sınıf sayısı (*m* = 2,…, 8).

g: Sınıflandırmada kullanılan özellik vektöründe yer alan finansal oran grubu

[*g* = 1 (sadece Likitide Oranları), 2 (sadece Mali Yapı Oranları), 3 (sadece Kaldıraç Oranları), 4 (sadece Piyasa Değeri Oranları), 5 (Hepsi)].

n: Özellik vektörlerinin geriye doğru göz önüne alındığı yılların sayısı (*n* = 1,…,5),

PKF algoritması temelli yaklaşımların performansı 10 kez çapraz doğrulama ile test edilmiştir. Performans göstergesi olarak, ortalama test başarısı olarak belirlenmiştir.

4.3 Hesapsal Sonuçlar

Farklı parametre değerleri için elde edilen test sonuçları Tablo 4.1’de verilmiştir.

Tablo 4.1 PKF algoritması temelli yöntemler ile on kez çapraz doğrulama test sonuçları

T	*m*	*g*	*N*	PKF Algoritması	Başarı Oranı
2003	2	5	1	1e1-PKF	0.78
2003	2	5	1	1eh-PKF	0.87
2003	3	5	1	1e1-PKF	0.75
2003	3	5	1	1eh-PKF	0.78
2003	6	5	1	1e1-PKF	0.62
2003	6	5	1	1eh-PKF	0.65
2003	8	5	1	1eh-PKF	0.45
2004	2	5	1	1eh-PKF	0.71
2004	2	5	2	1eh-PKF	0.67
2004	4	5	1	1eh-PKF	0.65
2004	4	5	2	1eh-PKF	0.66
2004	6	5	1	1eh-PKF	0.49
2004	6	5	2	1eh-PKF	0.55
2004	8	5	1	1eh-PKF	0.5
2004	8	5	2	1eh-PKF	0.49
2005	2	5	1	1eh-PKF	0.75
2005	2	5	2	1eh-PKF	0.76
2005	2	5	3	1eh-PKF	0.77
2005	4	5	1	1eh-PKF	0.75
2005	4	5	2	1eh-PKF	0.80
2005	4	5	3	1eh-PKF	0.72
2005	6	5	1	1eh-PKF	0.64
2005	6	5	2	1eh-PKF	0.65
2005	6	5	3	1eh-PKF	0.50
2005	8	5	1	1eh-PKF	0.50
2005	8	5	2	1eh-PKF	0.51
2005	8	5	3	1eh-PKF	0.40
2006	2	5	1	1eh-PKF	0.58
2006	2	5	4	1eh-PKF	0.55
2006	4	5	1	1eh-PKF	0.27
2006	4	5	3	1eh-PKF	0.31
2006	6	5	1	1eh-PKF	0.29
2006	6	5	4	1eh-PKF	0.26
2006	8	5	1	1eh-PKF	0.14
2006	8	5	3	1eh-PKF	0.14
2007	2	5	1	1eh-PKF	0.64
2007	2	5	3	1eh-PKF	0.57
2007	2	5	5	1eh-PKF	0.50
2007	4	5	1	1eh-PKF	0.53
2007	4	5	3	1eh-PKF	0.49
2007	4	5	5	1eh-PKF	0.42
2007	6	5	1	1eh-PKF	0.45
2007	6	5	5	1eh-PKF	0.30

2007	8	5	1	1eh-PKF	0.37
2007	8	5	5	1eh-PKF	0.28

Finansal oranların başarı oranı üzerinde etkisi Tablo 4.2'de verilmiştir.

Tablo 4.2 Finansal oranların 1eh-PKF'nin Başarı Oranı üzerinde etkisi

T	*M*	*g*	*n*	PKF Algoritması	Başarı Oranı
2003	2	1	1	1eh-PKF	0.87
2003	2	2	1	1eh-PKF	0.84
2003	2	3	1	1eh-PKF	0.84
2003	2	4	1	1eh-PKF	0.86
2003	2	5	1	1eh-PKF	0.87
2003	4	1	1	1eh-PKF	0.77
2003	4	2	1	1eh-PKF	0.77
2003	4	3	1	1eh-PKF	0.77
2003	4	4	1	1eh-PKF	0.76
2003	4	5	1	1eh-PKF	0.74
2003	8	1	1	1eh-PKF	0.49
2003	8	2	1	1eh-PKF	0.48
2003	8	3	1	1eh-PKF	0.48
2003	8	4	1	1eh-PKF	0.49
2003	8	5	1	1eh-PKF	0.45

Finansal oran gruplarının, Başarı Oranı üzerinde önemli bir etkisinin olmadığı görülmektedir. Başarı oranını etkileyen en önemli faktörler, grup sayısı ve şirketlerin gruplara dağılımıdır.

Geliştirilen Karar Destek Sistemi sayesinde seçilen finansal oranlara ve yıllara göre sınırsız sayıda sınıflandırma problemi oluşturmak mümkündür. Herhangi bir yıl için oluşturulacak sınıflandırıcı, o yılı kapsamayan geriye dönük belirli sayıda yıl için seçilen finansal oranları kullanmaktadır. Böylece elde edilen sınıflandırıcı, içinde bulunulan yıla ait finansal oranları kullanarak bir sonraki yıla yönelik tahmin yapmada kullanılabilmektedir. 2006 yılında şirketlerin getiri oranlarını tahmin etmek üzere bir yıl önceki, 2005 yılı, *Fiyat-Kazanç*, *Fiyat-Nakit Akışı*, *Piyasa Değeri- Defter Değeri*, *Sermaye Özsermaye*, *Uzun vadeli borçlar-Devamlı sermaye* finansal oranlarını kullanarak bir veri kümesi hazırlanmıştır. Hazırlanan veri kümesi için PKF algoritması ve tamsayılı yaklaşım kullanılarak sınıflandırıcı fonksiyonlar oluşturulmuştur. PKF algoritması ile elde edilen test başarısı %61.36, P matrisinin genişletilmesine izin verilen tek amaçlı tamsayılı yaklaşım ile ise test başarısı 0.32 tolerans değeri ile 18 fonksiyon kullanılarak oluşturulan sınıflandırıcı sayesinde %71.21 olarak elde edilmiştir. Bu problem ile ilgili farklı tolerans değerleri için yapılan analizler aşağıda verilen tabloda gösterilmektedir.

			0			0.3			0.32			0.35	
A0	fonk Say.	Ort_yk	Egitim_B	Test_B	Ort_yk	Egitim_B	Test_B	Ort_yk	Egitim_B	Test_B	Ort_yk	Egitim_B	Test_B
1	1	1	0.6389	0.5682	1	0.6541	0.5909	1	0.6541	0.6061	1	0.6541	0.5833
2	2	2	0.6869	0.5833	2	0.7062	0.6212	2	0.7062	0.6288	2	0.7071	0.6136
3	3	3	0.7248	0.6061	3	0.7424	0.6364	3	0.7433	0.6364	3	0.7441	0.6439
4	4	4	0.7534	0.6136	4	0.7753	0.6439	4	0.7753	0.6439	4	0.7786	0.6364
5	5	5	0.7778	0.6061	5	0.7997	0.6439	5	0.7997	0.6439	5	0.803	0.6515
6	6	6	0.798	0.6212	6	0.8216	0.6364	6	0.8216	0.6515	6	0.8249	0.6591
7	7	7	0.8157	0.6136	7	0.8392	0.6515	7	0.8392	0.6439	7	0.8434	0.6364
8	8	8	0.8325	0.6212	8	0.8561	0.6364	8	0.8561	0.6439	8	0.8611	0.6439
9	9	9	0.8485	0.6288	9	0.8712	0.6667	9	0.8712	0.6515	9	0.8771	0.6894
10	10	10	0.862	0.6288	10	0.8864	0.6742	10	0.8864	0.6667	10	0.8906	0.6818
11	11	11	0.8729	0.6212	11	0.8982	0.6591	11	0.899	0.6742	11	0.9015	0.6894
12	12	12	0.883	0.6439	12	0.9066	0.6667	12	0.9074	0.6742	12	0.9099	0.6742
13	13	13	0.8914	0.6439	13	0.915	0.6818	13	0.9158	0.6894	13	0.9184	0.697
14	14	14	0.8998	0.6136	14	0.9234	0.697	14	0.9242	0.697	14	0.9268	0.697
15	15	15	0.9082	0.6212	15	0.9318	0.6894	15	0.9327	0.697	15	0.9352	0.6894
16	16	16	0.9167	0.6136	16	0.9402	0.7045	16	0.9411	0.7045	16	0.9436	0.697
17	17	17	0.9251	0.6136	17	0.9487	0.6894	17	0.9495	0.697	17	0.952	0.7045
18	18	18	0.9335	0.6212	18	0.9571	0.7045	18	0.9579	0.7121	18	0.9604	0.7045
19	19	19	0.9419	0.6212	19	0.9655	0.697	19	0.9663	0.7045	19	0.9689	0.697
20	20	20	0.9503	0.6288	20	0.9739	0.697	20	0.9748	0.7045	20	0.9773	0.697
21	21	21	0.9588	0.6439	21	0.9823	0.7045	21	0.9832	0.7045	21	0.9857	0.7045
22	22	22	0.9672	0.6364	22	0.9891	0.7045	22	0.9899	0.7045	22	0.9924	0.7045
…	…	…	…	…	…	…	…	…	…	…	…	…	…
40	40	27	1	0.6439	24.4	1	0.6818	24.3	1	0.697	23.9	1	0.7121

Tek amaçlı tamsayılı yaklaşım ile elde edilen nihai sınıflandırıcı fonksiyon, 18 polihedral konik fonksiyonun noktasal minimumu olarak tanımlanmaktadır. 2005 yılının finansal oranlari kullanılarak 2006 yılını tahmin etmek üzere oluşturulan bu sınıflandırıcı fonksiyon, 2006 yılının finansal oranları kullanılarak 2007 yılı için şirketlerin hangi sınıfa ait olduğunu belirlemek üzere kullanılmıştır. Aşağıdaki tabloda g(x) sütunu bu fonksiyon ile elde edilen değerleri göstermektedir. 2007 yılındaki gerçek sınıf ile fonksiyon yardımıyla belirlenen sınıflar karşılaştırıldığında %66.91 sınıflandırma başarısı elde edilmiştir.

	K01	K02	G01	G02	G03	Sınıf	g(x)	Tahmin edilen sınıf
	1	2	3	4	5	6		
1	36.096	1.315	13.932	9.179	3.743	0	-0.79685	0
2	26.584	10.426	6.635	5.335	1.416	0	-0.61872	0
3	58.944	22.331	0	0	0.705	1	-0.12769	1
4	0.347	11.225	22.042	15.308	4.924	1	1.096358	0
5	7.71	44.409	0	0.956	0.338	1	0.542773	0
6	22.543	6.236	15.184	10.168	1.916	0	-0.94383	0

7	8.15	18.119	17.038	10.398	3.428	0	0.003545	1
8	4.395	16.796	28.639	3.607	0.559	0	-1.04969	0
9	28.674	12.657	16.978	6.77	0.863	1	-0.68499	1
10	6.894	5.008	14.166	9.362	1.082	0	-0.17405	0
11	61.799	7.309	0	18.866	0.975	1	0.810254	0
12	24.731	10.827	4.991	3.232	0.544	0	-0.99384	0
13	25.072	25.65	19.887	7.535	1.598	1	-0.21044	1
14	31.607	18.035	0	11.827	0.516	1	0.354976	0
15	9.151	7.424	11.516	7.765	1.818	1	-0.25683	1
.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.
122	25.576	17.208	17.039	10.229	2.778	0	-0.08168	0
123	92.957	8.243	0	21.271	1.329	1	1.155399	0
124	7.699	16.824	8.071	5.968	1.973	0	-0.75255	0
125	35.778	5.153	24.713	12.095	1.571	1	-0.16941	1
126	61.382	34.935	14.181	8.081	1.715	1	0.225417	0
127	47.787	20.041	12.445	5.832	2.374	0	-0.44769	0
128	33.814	2.975	5.517	4.796	0.901	0	-0.98524	0
129	15.076	36.763	9.32	4.833	0.787	1	-0.22774	1
130	112.619	25.142	9.793	4.495	1.42	1	-0.13689	1
131	36.181	24.122	0	40.932	0.792	1	3.363297	0
132	41.832	5.227	7.507	4.028	0.915	0	-0.85431	0
	5471.66	2463.101	2831.816	1563.829	223.725			
							Başarı	66.91%

Teşekkür

Projeye sağlamış olduğu desteğinden dolayı TÜBİTAK'a ve araştırma projelerinin desteklenmesi politikalarının oluşturulmasında emeği geçen herkese teşekkür ediyoruz. Projenin gerçekleştirilmesi sürecinde yakın desteğini esirgemeyen ve projeye ev sahipliği yapan İzmir Ekonomi Üniversitesine de teşekkür ediyoruz. Başından beri proje raporlarımızı titizlikle inceleyerek eleştirileri ile bu raporların güçlenmesine sağladıkları katkıdan dolayı danışman ve raportörlere de ayrıca teşekkür ediyoruz.

Sonuç ve Öneriler

Bu proje kapsamında, özellikle matematiksel programlama tabanlı yeni yaklaşımların geliştirilmesine odaklanılmıştır. Bu anlamda proje araştırmacılardan Gasimov ve Ozturk (2006) tarafından tanımlanan yeni bir fonksiyon sınıfı olan Polihedral Konik Fonksiyonlar çalışmanın merkezinde yer almıştır. Daha önce sunulmuş olan Polihedral Konik Fonksiyonlar Algoritmasının (PKFA) çalışma sistemi ve hem ikili hem de çoklu sınıflandırma problemleri üzerine uygulanmaları ile ilgili detaylı açıklamalar yapılmıştır. Polihedral konik fonksiyonlar algoritması ile sonlu adımda herhangi iki kümenin tam olarak ayrılabileceği ispatlanmış ve literatürdeki test problemleri üzerinde yapılan denemeler sonucunda test başarılarında oldukça iyi sonuçlar elde edilmiştir.

PKFA'nın eğitim ve test başarıları arasındaki farkının azaltılması, yani genelleştirme başarısının arttırılması düşüncesi, yeni yöntemlerin araştırılması için bir motivasyon sağlamıştır. Yine bu proje kapsamında tamsayılı programlama temeline dayanan iki amaçlı bir matematiksel model geliştirilmiştir. Bu iki amaçlı tamsayılı yaklaşım (IATS) ile sınıflandırıcı fonksiyonu oluşturacak PKF'ler, ayrılan nokta sayısını enbüyükleyecek ve seçilen fonksiyon sayısını enküçükleyecek şekilde bir araya getirilmektedir. Bu sayede genelleştirme başarısının arttığı gösterilmiştir.

Polihedral konik fonksiyonlar, doğrusal bir fonksiyonun l_1 normu ile genişletilmesi ile elde edilen bir fonksiyon sınıfı olarak tanımlanmıştır. Bu proje kapsamında l_2 ve l_{max} normları kullanılarak polihedral konik fonksiyonlar sınıfı genelleştirilmiştir. Bu sayede l_1 normu ile elde edilemeyen ayırıcı yüzeylerin oluşturulması sağlanmıştır. Gerek l_1 gerekse diğer normlar ile elde edilen fonksiyonların grafiklerinin nasıl oluştuğu, parametrelere karşı nasıl değiştiği ve ayırıcı yüzeyler arasındaki farklılıklar üç boyutlu grafikler yardımıyla detaylı şekilde gösterilmiştir. Yeni tanımlanan normlar ile PKFA'nın çok sınıflı problemler için başarıları da detaylı testler ile sunulmuştur.

Yine by projede şirket derecelendirme problemi, finansal oranlara dayalı bir sınıflandırma problemi olarak tariflenmiştir. Bu problemde gerçek sınıflar, belirli bir yıl için getiri oranlarından; özellik vektörleri ise getiri yılını kapsamayacak şekilde şirketlerin geçmişteki finansal oranlarından elde edilmiştir. Yüzlerce şirket, onlarca finansal oran için 2002-2007 yılları arasında bu şekilde neredeyse sınırsız sayıda sınıflandırma problemi tarif edilebilir. Bu sebeple, getiri oranları ve finansal oranların tutulması için bir veritabanı tasarlanmış ve bu veritabanı ile geliştirilen sınıflandırma yöntemlerinin bir araya getirildiği bir karar destek sistemi geliştirilmiştir. Böylece elde edilen veriler kullanılarak istenilen şekilde sınıflandırma problemlerinin oluşturulması sağlanmıştır.

Şirket derecelendirme probleminde iki çeşit doğrulama kontrolü gerçekleştirilmiştir. Birincisi, tanımlanan sınıflandırma probleminin 10 kez çapraz doğrulama sonucunda elde edilen doğru sınıflandırma oranıdır. İkincisi ise, özellik vektörünü oluşturan tüm finansal oranların yıl bazında bir dönem ileri kaydırılarak yeni özellik vektörlerinin elde edilmesi ile gerçekleştirilir. Bu özellik vektörlerinin bulunan sınıflandırıcı fonksiyonlar ile hesaplanması ve elde edilen sonuçların gerçek veriler ile karşılaştırılması ile de bir diğer başarı oranı hesaplanmıştır. Her iki yöntem ile elde başarı oranlarının birbirine yakın olması geleceğe yönelik olarak şirketlerin hangi getiri sınıfında yer alacağının belirlenmesi açısından önemli bir sonuçtur. Bu bilgi, tamamen karanlık bir ortamda bulunan yatırımcılara doğru kararı vermeleri yönünde ışık tutacaktır.

Bu proje kapsamında geliştirilen hem ikili hem de çoklu sınıflandırma yöntemlerinin performansı kıyaslanırken tüm dünyada yoğun kabul görmüş olan başarım ölçütlerinden doğru sınıflandırma oranı kullanılmıştır. Tüm doğru sınıflandırma oranları, test başarısının güvenilirliğini sağlamak adına 10 kez çapraz doğrulama yöntemi ile hesaplanmıştır. Bunun dışında k kez çapraz doğrulama yaklaşımının özel bir hali olan 5x2 çapraz doğrulama *F* testi, ikili sınıflandırmada kullanılan ROC (Receiver Operating Characteristics) eğrisi yöntemi veya bu eğri altındaki alanın ifadesi olan ve AUC (area under ROC curves) olarak da adlandırılan kıyaslama yöntemi ve benzeri diğer yöntemler elbette ki gelecekteki çalışmalarımızda bu çalışmaların doğası gereği uygun görülen yerlerde mutlaka kullanılacaktır.

Bu çalışmada çok farklı problemlerin çözümünde kullanılacak yeni sınıflandırma yöntemlerinin geliştirilmesi yanında şirket derecelendirme problemi özelinde bir uygulama gerçekleştirilmiştir. Burada gerçekleştirilen çalışmalar, konik fonksiyonlar temelli yeni yaklaşımların geliştirilebileceğine dair önemli ipuçları vermektedir. Gelecekte, yeni sınıflandırma yöntemleri üzerinde çalışılması; geliştirilen yöntemlerin ROC, AUC vb. gibi farklı doğruluk oranları ile de değerlendirilmesi; İMKB için geliştirilen yaklaşımın nihai bir ticari ürüne dönüştürülebilirliğinin ve farklı ülke borsaları için de bu çalışmanın uygulanabilirliğinin araştırılması planlanmaktadır.

Yayına gönderilen makaleler

1. A.M. Bagirov, J. Ugon, D. Webb, G. Ozturk and R. Kasimbeyli, Nonsmooth optimization in supervised data classification, (submitted to Springer Book).

2. A.M. Bagirov, J. Ugon, D. Webb, G. Ozturk and R. Kasimbeyli, A novel piecewise linear classifier based on polyhedral conic and max-min separabilities, (submitted to Mathematical Modelling and Numerical Optimization)

Hazırlanmakta olan makaleler:

1. A polyhedral conic functions based multi-objective integer programming approach to binary classification.
   İki sınıflı sınıflandırma için polyhedral konik fonksiyonlar temelli çok amaçlı tamsayılı programlama yaklaşımı.

2. A polyhedral conic functions based multi-objective integer programming approach to multi class classification.
   Çok sınıflı sınıflandırma için polyhedral konik fonksiyonlar temelli çok amaçlı tamsayılı programlama yaklaşımı.

3. The company rating problem for companies from İstanbul Stock Exchange and its solution using conic functions based classfication algorithms.
   İstanbul Menkul Kıymetler borsasındaki şirketler için derecelendirme problem ve konik fonksiyonlar temelli sınıflandırma algoritmaları ile çözümü.

Proje kapsamında sunulan konferans bildirileri:

1. A novel mathematical programming approach to classification based on linear and conic functions, 24th - European Conference on Operational Research, Lisbon, Portugal, July 11-15, 2010.

2. Kasimbeyli, R., Ozturk, G., Ustun, O., “Multi-class classification algorithms based on polyhedral conic functions and application to companies listed on the Istanbul Stock Exchange”, 14th International Congress on Computational and Applied Mathematics, Antalya, Turkey, 29 September - 2 October 2009.

3. Refail Kasimbeyli, “ A derivative free solution algorithm in nonconvex optimization”, 23rd European Conference on Operational Research, Bonn, Germany, July 5-8, 2009.

4. Ustun, O., Kasimbeyli, R., “Multiobjective portfolio optimization problem and solution methods”, 23rd European Conference on Operational Research, Bonn, Germany, July 5-8, 2009.

5. Ozturk, G., Kasimbeyli, R., “Conic functions in classification problems”, 23rd European Conference on Operational Research, Bonn, Germany, July 5-8, 2009.

6. Refail Kasimbeyli, “A cone separation theorem and scalarization in non-convex vector optimization”, 20th - EURO Mini Conference on Continuous Optimization and Knowledge-Based Technologies (EUROPT-2008), Neringa, Lithuania, May 20-23, 2008.

7. Ozturk G., Kasimbeyli R., Global tree optimization approachwith polyhedral conic functions and a multiobjective integer programming approximation to classification problems, EURO Mini Conference “Continuous Optimization and Knowledge-Based Technologies” EurOPT-2008, Neringa, Lithuania, May 20-23, 2008.

Kaynaklar

ABE, S., Inoue, T., Fuzzy Support Vector Machines for Multi-class Problems, Proc. European Symposium on Artificial Neural Networks, (2002) pp:113-118.

AKAL, Z., *İşletmelerde Performans Ölçüm ve Denetimi*, MPM Yayınları, Ankara, (2002).

AKGÜÇ, Ö., *Finansal Yönetim*, Avcıol Basın-Yayın, 6. Baskı, İstanbul, (1994).

AKTAŞ R., Doğanay M., Yıldız B., Mali başarısızlığın öngörülmesi: istatistiksel yöntemler ve yapay sinir ağı karşılaştırması, *Siyasal Bilgiler Fakültesi Dergisi*, 58 (4), (2003) s.1-24.

ALKAN A.L., Sanayi Şirketlerinin Performanslarının Finansal Göstergelerle Tahmini, *İMKB Dergisi*, Yıl 1, Sayı:4, Ekim- Aralık (1997).

ASTORINO A., Gaudioso M., Polyhedral separability through successive LP, *Journal of Optimization Theory and Applications*, 112 (2), (2002) pp. 265-293.

BAGIROV A.M., Max-min separability, *Optimization methods and software*, 20(2-3), (2005) pp. 271-290.

BAGIROV, A.M., Ugon, J., Supervised data classification via max-min separability, In: Applied Optimization, Vol.99: Continuous Optimization: Current Trends and Modern Applications, eds: Rubinov, A.M., Jeyakumar, V., Springer, (2005) pp.175-205.

BENNETT, K.P., Blue, J., A support vector machine approach to decision trees, Mathematics Report 97-100, Rensselaer Polytechnic Institute, Troy, NewYork, (1997).

BENNETT K.P., Mangasarian O.L., Robust linear programming discrimination of two linearly inseparable sets, *Optimization Methods and Software*, 1, (1992) pp.23-34.

BRADLEY P.S., Fayyad U.M., Mangasarian O.L., Data mining: overview and optimization opportunities, *INFORMS Journal on Computing*,11, (1999) pp. 217-238.

BREDENSTEINER E.J., Bennett K.P., Multicategory Classification by Support Vector Machines, *Computational Optimization and Applications*, 12, (1999) pp. 53–79.

CANBAŞ, S., Düzakın, H., Kılıç, S.B., Türkiye'de hisse senetlerinin değerlendirilmesinde temel finansal verilerin ve bazı makro ekonomik göstergelerin etkisi, Uludağ Üni., III. Ulusal Ekon. ve İst. Sem., Bursa, (1997).

CHEN, C., Mangasarian, O.L., Hybrid misclassification minimization. Mathematical Programming Technical Report, 95-05, University of Wisconsin, (1995).

CRAMMER K., Singer Y., On the Learnability and Design of Output Codes for Multiclass Problems, *Machine Learning*, 47, (2002) pp. 201-233.

DEMİR, A., Küçükkiremitçi, O., Pekkaya, S., Üreten, A., İMKB'deki Sanayi Şirketlerinin Hisse Senedi Getirileri ile Finansal Oranları Arasındaki İlişkilerin Belirlenmesi ve Bu İlişkilere Göre Şirketlerin Sıralandırılması (1992, 1993, 1994 Yılları İçin Bir Uygulama), SPK Yayın No: 56, (1997).

DUDA R., Hart P., *Pattern Classification and Scene Analysis*, New York: Wiley, (1973).

EASTON P., Harris T., Earnings as an explanatory variable for returns, *Journal of Accounting Research*, 29 (1), (1991) pp: 19-36.

EHRGOTT, M., *Multicriteria optimization*, 2nd Ed., Springer Verlag, (2005).

EHRGOTT, M., Waters, C., Gasimov, R.N., Ustun, O., Multiobjective Programming and Multiattribute Utility Functions in Portfolio Optimization, Technical Paper, The University of Auckland, New Zealand, (2006). http://www.esc.auckland.ac.nz/research/tech/esc-tr-639.pdf.

ERCAN, M.K., Ban, Ü., *Finansal Yönetim*, Gazi Kitabevi, Ankara, (2005).

FRIEDMAN, J.H., Another approach to polychotomous classificaion, Tech.Rep., (1996).

FUKUNAGA, K., *Introduction to Statistical Pattern Recognition*, 2nd ed., MA: Academic Press, Boston, (1990).

GASIMOV, R.N., Characterization of the Benson proper efficiency and scalarization in nonconvex vector optimization, Multiple Criteria Decision Making in the new millenium, Proceedings of the 15th International Conference on MCDM, Ankara, Turkey, July 10-14, 2000, Lecture Notes In Economics And Mathematical Systems, 507, Springer-Verlag Berlin Heidelberg, (2001) pp:189-198.

GASIMOV R.N., Ozturk G., Separation via Polyhedral Conic Functions, *Optimization methods and software*, 21(4), (2006) pp:527-540.

GASIMOV R.N., Sipahioglu A., Saraç T., A multi-objective programming approach to 1.5-dimensional assortment problem, *European Journal of Operational Research*, 179, (2007) pp:64-79.

GEOFFRION A.M., Proper efficiency and the theory of vector maximization, *Journal of Mathematical Analysis and Applications*, 22, (1968) pp:618–630.

HSU C.W., Lin C.J., A Comparison of Methods for Multiclass Support Vector Machines, *IEEE Transactions On Neural Networks*, 13(2), (2002) pp:415-425.

IBARAKI, T., Muroga, S., Adaptive linear classifiers by linear programming, Technical Report 284, Department of Computer Science, Univ. of Illinois, Urbana-Champaign, (1968).

JOHNSON, R.A., Wichern, D.W., *Applied Multivariate Statistical Analysis*, Prentice-Hall, Englewood Cliffs, NJ., (1992).

Kasimbeyli, R., A nonlinear cone separation theorem and scalarization in nonconvex vector optimization, *SIAM Journal on Optimization*, Volume 20, Issue 3, (2010) pp: 1591-1619.

KREBEL, U.H.G., Advances in kernel methods: support vector learning, Pairwise classification and support vector machines, Cambridge, Massachusetts: MIT Pres, (1999) pp. 255-268.

LEWANDOWSKI, A., Wierzbicki, A.P., Decision support systems using reference point optimization, Aspiration based decision support systems, Springer, Berlin, (1989).

MEGIDDO N., On the complexity of polyhedral separability, *Discrete and Computational Geometry*, 3, (1988) pp:325-337.

MICHIE, D., Spiegelhalter, D.J., Taylor, C.C., (Eds), *Machine Learning, Neural and Statistical Classification*, Ellis Horwood Series in Artificial Intelligence, London, (1994).

MIETTINEN K.M., Makela M.M., On scalarizing functions in multiobjective optimization, *OR Spectrum*, 24, (2002) pp:193–-213.

OU, G., Murphey, Y.L., Feldkamp, L., Multiclass pattern classification using neural networks, in ICPR '04: Proceedings of the Pattern Recognition, 17th International Conference on (ICPR'04) Volume 4, Washington, DC, USA, (2004) pp. 585-588.

ÖZDEMİR M.S., Gasimov R.N., The analytic hierarchy process and multiobjective 0–1 faculty course assignment, *European Journal of Operational Research*, 157 (2), (2004) pp:398–408.

ÖZER, G., Muhasebe karları ile hisse senedi verimleri arasındaki ilişkiler : İMKB'de deneysel bir analiz, SPK Yayın No:31, (1996).

QUILAN, J.R., C4.5: Programs for Machine Learning, San Mateo: Morgan Kaufmann, (1993).

SCHAOLKOPF, B., Smola, A. J., *Learning with kernels*, MIT Pres, Massachusetts, (2002).

STEUER R.E., *Multiple Criteria Optimization*: *Theory, Computation and Application*, Wiley, New York, (1986).

STRONG N., The relation between returns and earnings: evidence for the UK, *Accounting and Business Research*, 24 (93), (1993) pp:69-77.

WALTER, B.M., Robert, F.M., *Accounting: The Basis for Business Decisions*, McGraw-Hill, New York, (1988).

WANG Y-J, Lee H-S, A clustering method to identify representative financial ratios, *Information Sciences*, 178, (2008) pp:1087–1097.

WIERZBICKI, A.P., The use of reference objectives in multiobjective optimization, In: G. Fandel and T. Gal (Eds), MCDM theory and Application, Proceedings, Lecture notes in economics and mathematical systems, Springer Verlag: Hagen, 177, (1980) pp:468-486.

WIERZBICKI A.P., Reference point methods in vector optimization and decision support., Interim Report IR-98-017/ April, International Institute for Applied Systems Analysis, Laxenburg, Austria, (1998).

YALÇINER K., Atan M., Boztosun D., Finansal Oranlarla Hisse Senedi Getirileri Arasındaki İlişki, *Muhasebe Ve Finansman Dergisi*, 27, (2005) pp:176-187.

YU P.L., A class of solutions for group decision problems, *Management Science*, 936-949, (1973).

ZELENY M., Compromise programming. In: JL. Cochrane and M. Zeleny (Eds.), Multiple Criteria Decision Making, University of South Carolina, Columbia, (1973) pp: 262-301.

TÜBİTAK

PROJE ÖZET BİLGİ FORMU

Proje No	**:** 107M427
Proje Başlığı	**:** Veri madenciliğinde yeni çok-sınıflı sınıflandırma yöntemleri ve sirketlerin finansal özelliklerine göre derecelendirilmesi problemine uygulanması

Proje Yürütücüsü ve Araştırmacılar: Doç.Dr. Refail KASIMBEYLİ, Yrd. Doç.Dr. Gürkan Öztürk, Yrd. Doç.Dr. Özden Üstün

Projenin Yürütüldüğü Kuruluş ve Adresi: İzmir Ekonomi Üniversitesi

Destekleyen Kuruluş(ların) Adı ve Adresi:

Projenin Başlangıç ve Bitiş Tarihleri: 1 Şubat 2008 – 1 Şubat 2010

Öz (en çok 70 kelime)

Bu projede, iki ve daha fazla sınıfa sahip gerçek hayat veri kümelerinin sınıflandırılması problemi için matematiksel programlama temelli yeni yöntemler geliştirilmiş ve İstanbul Menkul Kıymetler Borsasında işlem gören şirketlerin derecelendirilmesi probleminin çözümü için uygulanmıştır. Literatürde sınıflandırma problemlerinin çözümünde kullanılan algoritmalarda şimdiye kadar sıklıkla kullanılan doğrusal fonksiyonlar yerine, projede sunulan bu yöntemlerde konik fonksiyonlar ve onların çeşitleri kullanılmıştır. Konik fonksiyonların yüksek performanslı ayırma kabiliyetine sahip yüzeyler oluşturduğu literatürde yaygın kabul görmüştür.

Anahtar Kelimeler: Veri madenciliği, iki-sınıflı sınıflandırma, çok sınıflı sınıflandırma, sınıflandırma algoritması, konik fonksiyonlar, tamsayılı programlama, matematiksel programlama, global optimizasyon, çok amaçlı optimizasyon, karar destek sistemi, finansal oranlar, şirketlerin derecelendirilmesi.

Fikri Ürün Bildirim Formu Sunuldu mu? **Evet** ☐ **Gerekli Değil** ☒

Fikri Ürün Bildirim Formu'nun tesliminden sonra 3 ay içerisinde patent başvurusu yapılmalıdır.

Projeden Yapılan Yayınlar: A.M. Bagirov, J. Ugon, D. Webb, G. Ozturk and R. Kasimbeyli, Nonsmooth optimization in supervised data classification, (submitted to Springer Book).