

KARAR AĞAÇLARI VE YAPAY SİNİR AĞLARININ SINIFLANDIRMA PERFORMANSLARININ KARŞILAŞTIRILMASI: AVRUPA BİRLİĞİ ÖRNEĞİ

Dilek ALTAŞ*
Vildan GÜLPİNAR**

ÖZET

Geleneksel istatistik teknikler sınıflandırma problemlerinde sıklıkla kullanılmaktadır. Ancak herhangi bir varsayıma sahip olmayan ve yapılan çok sayıda çalışma ile sınıflandırma performanslarının daha güçlü olduğu belirlenen Veri Madenciliği (VM) tekniklerine olan ilgi her geçen gün artmaktadır. Bu çalışmada, VM tekniklerinden Karar Ağaçları (KA) ve Yapay Sinir Ağları (YSA) teknikleri tanıtılarak, bu yöntemlerin sınıflandırma performansları bir örnek uygulama ile karşılaştırılmıştır.

Uygulamada Avrupa Birliği (AB)'ne üye ve aday ülkelere ait 9 makro ekonomik değişken verileri kullanılmıştır. Uygulamanın amacı bir ülkenin AB üyesi olmasında 'ekonomik gelişmişlik düzeyini arttırma' faktörünün etkili olup olmadığını tespit etmek ve etkili ise hangi değişkenlerin daha önemli olduğunu belirlemektir. Veriler, makine öğrenme paket programı olan WEKA yardımıyla analiz edilmiştir. KA analizi için C4.5. algoritması, YSA analizi için Çok Katmanlı Algılayıcılar yöntemi kullanılmıştır.

Anahtar Kelimeler: Karar ağaçları, Yapay sinir ağları, C4.5 Algoritması, Çok katmanlı algılayıcılar, Avrupa Birliği.

* Doç. Dr., Marmara Üniversitesi, İktisadi ve İdari Bilimler Fakültesi, Ekonometri Bölümü, İstatistik Bilim Dalı, mail: dilekaltas@gmail.com

** Marmara Üniversitesi, Sosyal Bilimler Enstitüsü, İstatistik Bilim Dalı Doktora Öğrencisi.

A COMPARISON OF CLASSIFICATION PERFORMANCES OF THE DECISION TREES AND THE ARTIFICIAL NEURAL NETWORKS: EUROPEAN UNION

ABSTRACT

The traditional statistics method is frequently used in classification problems. However data mining techniques draw more attention day by day because they do not have any assumption and their classification performance is determined to be stronger via many researches. In this study the Decision Trees and Artificial Neural Networks data mining techniques will be introduced and the classification performances of these methods will be compared with application sample.

In the application 9 macroeconomic variable data of member and candidate states of the European Union (EU) have been used. The aim of the application is to determine whether the factor of 'increasing the economical development level' is effective or not and if it is which variables are more important. The data have been analyzed via machine training package program of WEKA. C4.5 algorithm has been used for KA analysis whereas Multilayer Perceptrons have been used for YSA analysis.

Key Words: *Decision trees, Artificial neural networks, C4.5 Algorithms, Multilayer perceptrons, European Union.*

1. GİRİŞ

Veri Madenciliği (VM), verilerde daha önce bilinmeyen gizli örüntüleri açığa çıkaran ve bu bilgilerden yararlı sonuçlar elde ederek geleceğe yönelik tahmin yapmaya imkân tanıyan dinamik bir süreçtir. Son yıllarda doğrusal olmayan ilişkileri de analiz edebilen ve geleneksel istatistiksel yöntemler gibi varsayımlara ihtiyaç duymayan VM teknikleri yoğun ilgi görmektedir.

Karar Ağacı (KA) yöntemi, kolay anlaşılabilir kurallar üretmesi, kuralların görselleştirilebilmesi ve farklı değişken yapılarına uygun güçlü algoritmalara sahip olması sebebiyle en sık kullanılan VM tekniklerindedir. Yapay Sinir Ağları (YSA) ise doğrusal olmayan, çok boyutlu, gürültülü, karmaşık, eksik ve kusurlu verilerin olması ve problemin çözümü için özellikle bir matematiksel modelin ve algoritmanın bulunmaması hallerinde de kullanılabilirler için bilimin her alanında sıklıkla kullanılan VM tekniklerindedir.

Günümüzde AB, AB üye ülkeleri ve Türkiye ilişkilerini analiz eden pek çok araştırma yapılmış, özellikle AB ülkelerinin makro ekonomik

göstergeleriyle Türkiye'ye ait verilerin karşılaştırılması bir çok çalışmanın konusu olmuştur.

Beş dönemden oluşan AB genişleme süreci şu şekildedir: 1. genişleme İngiltere, İrlanda ve Danimarka'nın; 2. genişleme Yunanistan, İspanya ve Portekiz'in; 3. genişleme İsveç, Finlandiya ve Avusturya'nın; 4. genişleme Kıbrıs Rum Kesimi, Polonya, Macaristan, Çek Cumhuriyeti, Slovenya, Estonya, Letonya, Litvanya, Slovakya ve Malta'nın ve 5. genişleme Bulgaristan ve Romanya'nın AB'ye katılımı ile gerçekleşmiştir. Diğer AB üyesi ülkelerle karşılaştığımızda tam üyelik müracaatından sonra en uzun süreçte değerlendirilen ülke Türkiye olmuştur ve halen tam üyelik müzakereleri sonuçlandırılmış değildir. "Türkiye ne yapmalı?" sorusuna cevap ararken, öncelikle "AB, üyelik için hangi kıstasların yerine getirilmesini bekliyor ve yerine getirilmesi gereken kıstasların hangileri adaylık sürecini ne derecede etkiliyor?" sorusu sorulmalıdır. Aday ülkenin ne kadar kısa sürede "üye ülke" kabul edileceği, bu kıstasların yerine getirilme hızıyla ilgilidir. Bu çalışmada aday ülkelerin müzakere sürecinin¹ uzunluğu ile ekonomik verilerinde görülen değişimin ilişkili olup olmadığı analiz edilecektir.

AB'nin genişleme sürecinin üye ülkeler üzerindeki etkilerini makro ve mikro değişkenler aracılığıyla inceleyen çalışmalar Breuss (2002)² ve Doyle ve Fidrmuc (2003)³ tarafından yapılmıştır. Brada ve Kutan (2001)⁴ tarafından yapılan çalışmada Almanya'nın para politikası ile geçiş ekonomisine sahip AB aday ülkeleri, geçiş ekonomisine sahip olmayan AB aday ülkeleri ve yeni üye olmuş ülkelerin yakınsamaları karşılaştırılmış, aday ülkelerin makro ekonomik gelişmişliklerinin üyelik sürecine olumlu

¹Veriler derlenirken 'müzakere başlangıç tarihi' değil, daha gerçekçi sonuçlar elde etmek için 'başvuru tarihi' dikkate alınmıştır. Literatürde bu ayrım gözetilmeden 'müzakere süreci' kavramı kullanıldığı için çalışma kapsamında AB'ye aday ülkelerin başvuru yaptıkları yıl ile aday kabul edildikleri yıl arasındaki süreç "müzakere süreci" ile ifade edilecektir.

² Fritz Breuss, 'Benefits and Dangers of EU Enlargement', *Empirica, Kluwer Academic Publishers*, Vol 29, Netherlands 2002, s.245-274.

³ Orla Doyle-Jan Fidrmuc, 'Who is in Favor of Enlargement? Determinants of Support for EU Membership in the Candidate Countries' Referanda', <http://www.fidrmuc.net/research/refd.pdf>, (09.10.2011).

⁴ J.C. Brada-A.M. Kutan, 'The Convergence of Monetary Policy Between Candidate Countries and the European Union', *Economic Systems*, Vol. 25, No. 3, 2001, s. 215-231.

etkisi üzerinde durulmuştur. Halkos ve Tzeremes (2009)⁵ tarafından yapılan çalışmada 25 üye ülkenin 1995-2005 yılları arasında kaydettikleri ekonomik ilerleme değerlendirilmiş, eski 15 üye ülkenin AB genişlemesinden ne şekilde etkilendiği analiz edilmeye çalışılmıştır. Çalışmada üye ülkelerin aday ülke olabilmelerinde ekonomik gelişmişliğin önemli ancak yeterli olmadığı sonucuna varılmıştır.

AB üye ve aday ülkelerinin makro ekonomik ve demografik değişkenlerinin karşılaştırıldığı bu çalışmada, ortaya çıkan sınıflandırma problemi, sınıflandırma başarılarının geleneksel istatistiksel yöntemlere göre üstünlüğü birçok çalışma ile gösterilmiş olan KA ve YSA yöntemleriyle analiz edilecek ve sonuçlar tartışılacaktır.

2. VERİ VE YÖNTEM

2.1. Karar Ağaçları

Karar Ağaçları, istatistiksel olarak anlamlı grupları bulan ve cevapları açık bir şekilde, kolay okunabilir ağaç diyagramları ile veren, gözlemleri sınıflayan ya da tahmin eden kurallar grubudur.⁶ KA, bağımlı değişken üzerindeki farklılıkların maksimize edilmesi amacıyla veri setinin sıralı bir şekilde bölünmesini ifade eder.⁷ Verileri belli değişken değerlerine göre sınıflandırmaya yarayan KA'da kullanılan algoritmalarda girdiler ve çıktılar verilerin belirlenen değişkenleridir ve KA algoritması çıktı veri değişkenleri için girdi veri değişkenlerini veri yapıları ile keşfeder.⁸

KA, etkili bir sınıflandırma yöntemidir ve KA sunumu, en yaygın kullanılan mantık yöntemidir. Esas olarak makine öğreniminde ve uygulamalı istatistik literatüründe tanımlanan çok fazla sayıda KA tümevarım algoritması vardır. Bu algoritmalar, bir seri girdi-çıkı kümesinden KA oluşturan denetlemeli öğrenme yöntemleridir. Tipik bir KA öğrenme sistemi, araştırma alanının bir kısmında çözüm arayan, 'yukarıdan

⁵ G. Halkos-E. Tzeremes, 'Economic Efficiency and Growth in the EU Enlargement', *Journal of Policy Modeling*, Vol. 31, 2009, s.842-862.

⁶ Nurhan Doğan-Kazım Özdamar, 'Chaid Analizi ve Aile Planlaması ile Bir Uygulama', *T. Klin Tıp Bilimleri Dergisi*, Cilt: 23, Eskişehir 2003, s.392-397.

⁷ Doruk Dorman, Bankacılık Sektöründe Müşteri İlişki Yöntemi: CRM Açısından Veri Madenciliği Yöntemi, Doktora tezi, İstanbul 2003, s.34.

⁸ P.N. Tan-M. Steinbach-V. Kumar, *Introduction to Data Mining*, Pearson Addison Wesley, Boston 2006, s.146.

aşağıya' yöntemini benimser. Bu yöntem basit bir ağacın (en basiti olması şart değil) bulunabileceğini garanti eder. Bir KA, değişkenlerin test edildiği yerlerde düğümler içerir. Bir düğümden dışa açılan dallar, düğümdeki testin bütün olası sonuçlarına karşılık verir.⁹

KA endüksiyonu, sınıflandırma modellerinin oluşturulmasına yönelik parametrik olmayan bir yaklaşımdır. KA algoritmaları aşırı donatımdan kaçınma yöntemlerinin uygulanması esnasında gürültü (*noisy*) oluşmalarına karşı oldukça dirençlidirler. Gereğinden fazla değişkenlerin varlığı, KA'nın doğruluğunu olumsuz yönde etkilememektedir. KA oluşturmak için geliştirilen teknikler, hesap olarak oldukça ucuzdur ve düzenleme boyutu çok geniş olsa da hızlıca model oluşturmayı mümkün hale getirmektedirler.¹⁰

KA'nın sağladığı avantajlar yanında dezavantajları da vardır. Genellikle karar vermek KA yaklaşımı için ciddi bir sorun oluşturur. Sorun şundan kaynaklanır: Ağaç, ayrılan bölümleriyle genişledikçe, yapılan sınıflandırma sonucunda bölüm düğümlerinde daha az bilgi kalacaktır. KA veriyi çok fazla sayıda takıma böler. Bu takımlar daha spesifik bir hal aldıkça küçülmeye başlar. İncelenmesi gereken farklı durumların sayısı attıkça, eğitim kümelerinin her biri daha da küçülür.

Rakamlardaki azalma yüzünden sınıflandırmanın doğru bir şekilde gösterilmesinde daha az güvenilirlik söz konusu olur. KA çok fazla sayıda küçük dallardan oluşturulursa bu düğümlerin içerisinde herhangi doğru olabilecek bir istatistiksel incelemeden başarıyla çıkabilecek kuralların bulunması iyimser bir olasılıktır. Çünkü genellikle bu dallardan çıkacak olan her bir düğüm, olası bütün sınıfların küçük oranlarını içerecektir. Bu da uygulamada sorunlara yol açabilir.¹¹

KA dinamik bir veriyi kolayca ele alamazlar. Bu nedenle değişkenin etki kümeleri kolayca incelenecek şekilde kategorilere bölünmelidir.¹² KA her dalda ancak belli bir özellik açısından veri kümelerini birbiriyle karşılaştırabilmektedir. Bu nedenle birden fazla değişken söz konusu olduğunda değişkenler arasındaki ilişkiyi saptama konusunda yetersiz kalmaktadır.

⁹ Mehmed Kandartzic, *Data Mining Concepts Models and Algorithms*, John Wiley&Sons, 2003 (*e-book, part 7*).

¹⁰ Tan ve diğerleri, a.g.e., s.152.

¹¹ Claude Seidman, *Data Mining with Microsoft SQL Server 2000*, Microsoft Press, Washington 2001, s.69.

¹² M. H. Dunham, *Data Mining Introductory and Advanced Topics*, Pearson Education Inc., Southern Methodist 2003, s.92.

KA modelleri sayısal değerleri kategorik olarak değerlendirmektedir. Ayrıca kategorilerde yer alan veri sayısı azaldıkça KA'nın hatalı sonuç verme eğilimi artmaktadır.

Verilerin karmaşıklığı arttıkça KA'da oluşan dalların sayısı da artmakta ve bu durum KA'yı yönetmeyi zorlaştırmaktadır.

Bir alt KA, bir KA'da çok kez tekrarlanabilmektedir. Bu ise, KA'yı, gereğinden daha karmaşık hale getirmekte ve yorumlanmasını zorlaştırmaktadır. Böyle bir durum her bir dahili düğümde tek bir değişken testi koşuluna dayanan KA uygulamalarından doğabilir. KA algoritmalarının çoğu bir 'böl ve yönet' bölünme stratejisi kullandığı için aynı test koşulu, değişken alanındaki farklı bölümlere uygulanabilir ve bu da alt ağaçların tekrarlanması (çoğalması) problemini beraberinde getirir.¹³

2.1.1. C 4.5 Algoritması

C4.5 algoritması, KA yönteminde en sık kullanılan algoritmalarından biridir.

C4.5 algoritmasının genel özellikleri şu şekilde sıralanabilir:¹⁴

- *Eksik veri*: KA kurulduğunda eksik veri basitçe ihmal edilir. Yani kazanç oranı yalnızca, söz konusu parametre için bir değere sahip diğer kayıtlara bakılarak hesaplanır. Eksik bir parametre değeri olan bir kaydı sınıflandırmak için söz konusu kalemin değeri, diğer kayıtlar için parametre değerlerine ilişkin bilgiler kullanılarak tahmin edilebilir.

- *Sürekli veri*: Sayısal değerlere sahip değişken içerisinde uygun eşik değeri bulunduktan sonra ikili ya da daha çok bölünme ile veri kümesi bölünebilir.

- *Budama*: C4.5'te önerilen iki ana budama stratejisi mevcuttur:

- 'Alt ağaç değiştirmeli' adı verilen stratejide değiştirme sonucunda, başlangıçtaki ağacınkine yakın hata oranı elde edilebiliyorsa alt ağaç, yaprak düğümlerle değiştirilir. Alt ağaç değiştirme, ağacın altından yukarıdaki köküne kadar uygulanır.

- 'Alt ağaç yükseltme' adı verilen diğer budama stratejisiyle alt ağaç, en çok kullanılan kendi alt ağacıyla değiştirilir. Böylece alt ağaç, mevcut konumundan ağacın daha üst noktasındaki düğüme yükseltilir. Yine de bu değiştirme için hata oranındaki artışı tespit etmemiz gerekmektedir.

¹³ Tan ve diğerleri, a.g.e., s. 152.

¹⁴ Dunham, a.g.e., s.100.

• *Ayırma*: Aşırı uyum (*overfitting*) sebebiyle oluşan hata, C4.5 tarafından geliştirilen yöntemle telafi edilmeye çalışılmaktadır.

2.2. Yapay Sinir Ağları

YSA, insanlar tarafından gerçekleştirilmiş örnekleri (gerçek beyin fonksiyonlarının ürünü olan örnekleri) kullanarak olayları öğrenebilen, çevreden gelen olaylara karşı nasıl tepkiler üretileceğini belirleyebilen bilgisayar sistemleridir. İnsan beyninin fonksiyonel özelliklerine benzer şekilde, öğrenme, ilişkilendirme, sınıflandırma, genelleme, özellik belirleme, optimizasyon gibi konularda başarılı bir şekilde uygulanmaktadırlar. YSA, örneklerden elde ettikleri bilgiler ile kendi deneyimlerini oluşturur ve daha sonra benzer konularda benzer kararları verirler.

YSA birbirine hiyerarşik olarak bağlı ve paralel olarak çalışabilen yapay hücrelerden oluşmaktadır. Bir yapay sinir ağının en temel görevi, kendisine gösterilen bir-girdi setine karşılık gelebilecek bir çıktı seti belirlemektir. Bunu yapabilmek için ağ, ilgili olayın örnekleri ile eğitilerek (öğrenme) genelleme yapabilecek yeteneğe kavuşturulur. Bu genelleme ile benzer olaylara karşılık gelen çıktı setleri belirlenir. Özellikle olaylar hakkında bilgilerin olmadığı fakat örneklerin bulunduğu durumlarda çok etkin olarak kullanılacak bir karar verme aracı ve hesaplama yöntemi olarak görülebilir.¹⁵

YSA, doğrusal olmayan, çok boyutlu, gürültülü, karmaşık, kesin olmayan, eksik verilerinin olması ve problemin çözümü için özellikle bir matematiksel modelin ve algoritmanın bulunmaması hallerinde yaygın olarak kullanıldıkları görülmektedir. YSA makine öğrenmesi gerçekleştirirler. Diğer programlarda olduğu gibi veriler bir veri tabanında veya programın içinde gömülü değildir. Bilgiler ağın üzerinde saklı olup ortaya çıkartılması ve yorumlanması zordur. Diğer taraftan YSA'nın güvenle çalıştırılabilmesi için önce eğitilmeleri ve performanslarının test edilmesi gerekmektedir. Veri girişi yapılırken sadece nümerik bilgiler ile çalışabilmektedirler. YSA'nın, kendi kendini organize etme ve öğrenebilme yetenekleri vardır. Önemli avantajlarından biri de belirsiz, tam olmayan bilgileri işleyebilme özellikleridir. En önemli dezavantajları ise probleme uygun ağ yapısının belirlenmesinin genellikle deneme yanılma yolu ile yapılması dolayısıyla en iyi çözümün garanti edilememesi ve ağ

¹⁵ Ercan Öztemel, *Yapay Sinir Ağları*, Papatya Yayıncılık, İstanbul 2006, s.29-30.

davranışlarının açıklanamamasıdır.¹⁶¹⁷

Yapay sinir hücreleri bir araya gelerek YSA'yı oluştururlar. Sinir hücrelerinin bir araya gelmesi rasgele olmaz. Genel olarak hücreler 3 katman halinde ve her katman içinde paralel olarak bir araya gelerek ağı oluştururlar. Bu katmanlar:¹⁸

- Girdi katmanı: Bu katmandaki proses elemanları dış dünyadan bilgileri olarak ara katmanlara transfer etmekle sorumludurlar. Bazı ağlarda girdi katmanında herhangi bir bilgi işleme olmaz.
- Ara katmanlar: Girdi katmanından gelen bilgiler işlenerek çıktı katmanına gönderirler. Bu bilgilerin işlenmesi ara katmanlarda gerçekleştirilir. Bir ağ için birden fazla ara katman olabilir.
- Çıktı katmanı: Bu katmandaki proses elemanları ara katmandan gelen bilgileri işleyerek ağın girdi katmanından sunulan girdi seti (örnek) için üretmesi gereken çıktıyı üretirler. Üretilen çıktı dış dünyaya gönderilir.

Girdi ve çıktı katmanları arasında deneme ve yanılma ile optimize edilen belli sayıda nöron ile birlikte gizli bir katman vardır. Tek bir katman hesaplama zamanını kısaltır ve genelde gizli katmanları olan YSA'lar ile aynı sonucu verir.¹⁹ Gizli nöronların sayısı ağ sapması ve varyansı arasındaki alış verişini minimize etmek için seçilir.²⁰

Bazı ağlarda ağı parametre değerlerinin (mesela öğrenme katsayısı, her katmanda olması gereken proses elemanı (yapay hücrelerin) sayısı, katman sayısı vb. belirlenmesinde de bir kural olmaması diğer bir problemdir. Bu parametrelerin belirlenmesi de kullanıcının tecrübesine bağlıdır.

2.2.1. Çok Katmanlı Algılayıcılar

YSA modelleri içerisinde en çok kullanılan ağ tipidir. Bu ağlarda girdi sinyali katmanlar boyunca ileriye doğru işlenerek çıktı değeri üretilmektedir. Bu tip bir ağda bilgi sırasıyla üç katmandan ileri doğru geçirilmektedir. İlk katman olan girdi katmanının görevi bilgiyi bir sonraki katman olan gizli katmana aktarmaktır. Burada veri herhangi bir işlemde geçirilmemektedir.

¹⁶ Simon Haykin, *Neural Networks and Learning Machines*, Third edition, Pearson Education, 2009, s.34-35.

¹⁷ Öztemel, a.g.e., s.31-33.

¹⁸ Öztemel, a.g.e., s.52-53.

¹⁹ C. M. Bishop, *Neural Networks for Pattern Recognition*, Oxford University Press, London 1995, s. 192.

²⁰ Bishop, a.g.e., s.340.

Gizli katman olarak adlandırılan ikinci katmanda yer alan sinir hücreleri veriyi transfer fonksiyonundan geçirerek bir sonraki katman olan çıktı katmanına yollarlar. Çıktı katmanı kendisine gelen bu bilgiyi kendi transfer fonksiyonundan geçirerek çıktıyı üretmiş olur. Elde edilen çıktı istenilen çıktı değeriyle karşılaştırılır. Ortaya çıkan hata payı ağ boyunca ağırlıklara geri yayılır. Düzenlemeler yapılarak veri seti ağa tekrar sunulur, başta anlatılan süreç tekrarlanarak çıktı tekrar elde edilir ve istenilen çıktı değeriyle karşılaştırılır. Bu işlem istenilen çıktıya kabul edilir hata değerinca yaklaşılan kadar sürdürülür. Bu işlem için kullanılan en yaygın kabul görmüş algoritma geri yayılım algoritmasıdır.

ÇKA ağları denetimli öğrenme stratejisine göre çalışırlar. Şöyle ki, bu ağlara eğitim sırasında hem girdiler hem de o girdilere karşılık üretilmesi gereken (beklenen) çıktılar gösterilir. Ağın görevi her girdi için o girdiye karşılık gelen çıktıyı üretmektir. ÇKA ağının öğrenme kuralı en küçük kareler yöntemine dayalı Delta Öğrenme Kuralının genelleştirilmiş halidir. O nedenle öğrenme kuralına *Genelleştirilmiş Delta Kuralı* da denmektedir. Verilerin bir kısmı ağ eğitmek için kullanılırken, diğer kısmı test etmek için kullanılır. Delta Kuralına göre önce ağın çıktısı hesaplanır daha sonra da hata oranını min. yapacak şekilde ağırlıklar değiştirilir.

2.3. Karar Ağaçları ve Yapay Sinir Ağları Yöntemlerinin Karşılaştırılması

KA ve YSA değişkenler arasında doğrusal olmayan ilişki bulunan verileri modelleyebilmekte ve her iki yöntem de değişkenler arası etkileşimi yönetebilmektedirler. Ancak aralarında belli başlı farklılıklar vardır. YSA anlaşılabilir bir model sunmaz. Buna karşılık bir KA incelendiğinde, başlangıç değişkeninin veriyi iki kategoriye böldüğünü ve daha sonra diğer değişkenlerin sonuç olarak çıkan grupları böldüğünü görmek mümkündür. Bu bilgi analiz edilen verinin arkasındaki yapıyı anlamayı kolaylaştırır. Ancak YSA için bunu söylemek mümkün olmaz.²¹

KA ve YSA'nın modellemede değişkenlerin önemini anlama konusunda zıt bir ilişkisi vardır. KA modelinin yapı ve biçimi, araştırmacıların her bir girdinin ağacın yapısına nasıl katkıda bulunduğunu anlamalarına izin verirken, YSA bu kadar kolay yorumlanamaz. Bu yüzden,

²¹ G. W. H. German-G. West-M. Gahegang, Statistical and AI Techniques in GIS Clasification: A Comparison, http://divcom.otago.ac.nz/Sirc/conferences/1999/14_German.pdf, (11.10.2011).

YSA yaklaşımı kullanılırken, ortaya çıkan modelin değişken seçimi ve hassasiyet analizi, farklı girdilerin önemi hakkında bilgi sahibi olabilmek açısından önemlidir.²²

Yukarıdaki açıklamalar ışığında KA'nın, YSA'dan farklı olarak, kolaylıkla eğitilebildiği, çok hızlı bir şekilde doğru sonuçlar verdiği ve her adımının kolaylıkla izlenilip anlaşılabilirliği söylenebilir. KA'nın geleneksel istatistiksel sınıflayıcılardan farkı ise, kayıp ve gürültülü veriyle uyumlu çalışabilmesi ve parametresiz bir sınıflayıcı olmasıdır.²³ Parametresiz sınıflayıcılar, istatistiği temel almazlar, bu nedenle, verinin özelliklerinden bağımsızdırlar ve eğitici kümenin dağılımını hesaba katmazlar. KA ve YSA en çok bilinen parametresiz sınıflayıcılarıdır.

Uygulamaya geçmeden önce KA ve YSA'nın çalışma prosedürlerini karşılaştırmak şüphesiz uygulamanın anlaşılmasını kolaylaştıracaktır.

KA'nın çalışma prosedürü şu şekildedir.²⁴

KA Veri kümesi P 'nin eğitim değişkenleri (bağımsız değişkenler) grubundan bir karar ağacı oluşturulur.

Girdi: Girdi, veri kümesi P ile gösterilir.

Çıktı: Çıktı, bir karar ağacıdır.

Adımlar:

- (1) Bir N düğümü oluşturulur.
- (2) Eğer P 'deki değişkenler grubunun hepsi aynı C sınıfı ise; N 'ye C sınıfı etiketli bir yaprak düğümü olarak dönülür.
- (3) Eğer aday değişkenlerin bir kümesi yok ise, o zaman N 'ye çoğunluk sınıfı P ile etiketli bir yaprak düğümü olarak dönülür.
- (4) En iyi ayrılma noktasını bulmak için bölünme kriterinin belirlendiği bir prosedür (eşik değer) uygulanır.
- (5) N düğümü, en iyi ayrılma noktası ile etiketlenir.
- (7) Ayrılma noktasının her bir j çıktısı için değişkenler kümesi bölünür ve her bir kısım için alt-ağaçlar geliştirilir.

²² Tina Tireli-Daniela Pessani, 'Use of Decision Tree and Artificial Neural Network Approaches to Model Presence/Absence of *Telestes Muticellus* in Piedmont', *Rever Research and Applications*, Vol. 25, Torino 2009, s. 1008.

²³ German ve diğerleri, a.g.m., s.7

²⁴ P. Kumar-Nitin-V. Sehgal -D.S. Chauhan, 'Performans Evaluation of Decision Tree versus Artificial Neural Network Based Classifiers in Diversity of Datasets', *World Congress on Information and Communication Technologies(WICT 2011)*, Mumbai 2011, s. 799.

(8) P^j 'yi doğrulayan j çıktısı için P^j bir veri değişkenler kümesini oluşturur.

YSA'nın çalışma prosedürü şu şekildedir.²⁵

Geri yayılım Algoritması

Geri yayılım algoritması kullanarak sınıflandırma için sinir ağı öğrenme basamakları aşağıda gösterilmiştir.

Girdi: Girdi, örneklerdir. Öğrenme oranı l ile gösterilmektedir ve Çok Katmanlı İleri Beslemeli bir ağ oluşturulur.

Çıktı: Çıktı, örnekleri sınıflandırmak üzere eğitilmiş bir sinir ağıdır.

Adımlar:

(1) Ağdaki tüm ağırlıklar (w) ve eşik değerleri (θ) başlatılır.

(2) Girdinin ileri doğru yayılması (girdi katmanından işlenerek çıktı katmanına doğru).

(3) Her bir gizli katmanda veya çıktı katmanında birim j

$$I_j = \sum w_{ij} O_i + \theta_j \text{ önceki katmana göre her bir } j \text{ biriminin net çıktısı}$$

hesaplanır.

$$(4) O_j = \frac{1}{1 + e^{-I_j}} \text{ her bir } j \text{ biriminin çıktısı hesaplanır.}$$

(5) Hataların geriye doğru yayılması.

$$(6) Err_j = O_j (1 - O_j) (T_j - O_j) \text{ hata hesaplanır.}$$

(7) En sondan ilk gizli katmana kadar gizli katmanlardaki her bir j birimi için $Err_j = O_j (1 - O_j) \sum Err_k W_{jk}$ bir sonraki daha yüksek katmana k^j ya göre hata hesaplanır.

(8) Ağdaki her bir W_{ij} ağırlığı için

$$\Delta W_{ij} = (l) Err_j O_i \text{ ağırlık artışı}$$

$$(9) W_{ij} = W_{ij} + \Delta W_{ij} \text{ ağırlık güncellemesi}$$

(10) ağdaki her bir θ_j eşik değeri için

$$\Delta \theta_j = (l) Err_j \text{ eşik değer artışı}$$

$$(11) \theta_j = \theta_j + \Delta \theta_j \text{ eşik değer güncellemesi yapılarak}$$

sonuçlandırılır.

²⁵ Kumar ve diğerleri, a.g.m., s. 800.

3. UYGULAMA

Yukarıda anlatılan KA ve YSA yöntemlerinin sınıflandırma performansları AB örneği yardımıyla karşılaştırılacaktır.

Uygulamanın amacı AB üyelik sürecinde ekonomik gelişmişlik kıstasının müzakere sürecine etki edip etmediği ve eğer etkili ise bu süreçte hangi ekonomik kıstasların daha belirleyici olduğunu ortaya çıkarmaktır. Ayrıca çalışma sonucunda oluşturulan model, AB'ye tam üyelik başvurusunda bulunacak ülkelerin, müzakere sürecinin kaç yıl süreceğini tahmin etmelerine olanak sağlayacaktır.

Bu amaç doğrultusunda AB üyesi 20 ülke ve Türkiye'nin makro ekonomik verilerinden enflasyon oranları, kur oranları, işsizlik oranları, ihracat, ithalat, toplam iş gücü, sabit sermaye yatırımları, gayri safi yurtiçi hâsıla (GSYİH) ve nüfus yoğunluğu değişkenleri incelenmiştir. Makro ekonomik veriler, ülkelerin tam üyelik başvurusu yaptıkları yıl ile üye kabul edildikleri yılda görülen değerlerin değişim miktarları alınarak hesaplanmıştır. Kurucu ülkelerin adaylık süreci olmadığından çalışmanın kapsamı dışında tutulmuştur.

Bu değişkenlerin seçilmesinde Türkiye için 1998 yılından bu yana her yıl yayınlanan ve Türkiye'nin AB müktesebatına uygunluğunun değerlendirildiği ilerleme raporları (2002-2010) ve Avrupa Topluluğu istatistik ofisi (*Eurostat*) makro ekonomik göstergeler çizelgesi baz alınmıştır.²⁶

Çalışmada veri seti şu şekilde elde edilmiştir:

- (1) Uygulama kapsamına alınan tüm ülkelerin AB'ye tam adaylık başvurusu yaptığı tarih ve üye olduğu tarih tespit edilmiştir.
- (2) Ülkelerin tam adaylık başvurusundan itibaren ortalama kaç yıl sonra üye oldukları tespit edilmiştir (ortalama 9 yıl). Buna göre 9 yıldan daha az bekleyen ülkeler 1, 9 yıldan daha fazla bekleyen ülkeler ise 2 sınıf değeriyle etiketlenmiştir.
- (3) Aynı işlem bağımsız değişkenler için de tekrarlanmıştır. Örneğin, Macaristan 1994 yılında tam adaylık başvurusu yapmış ve 2004 yılında üye sıfatı kazanmıştır. Değerlendirme süreci 10 yıl olduğu için Macaristan'ın sınıf etiketi 2'dir. Aynı şekilde ülkelere ait makro ekonomik değişkenlerde para birimi farklılıkları ortadan kaldırılarak tüm ülkelerin ortalamasının altında ya da üstünde bir

²⁶http://ec.europa.eu/economy_finance/indicators/annual_macro_economic_database/ameco_applet.htm

değer seyretmesine göre sırasıyla 1 ve 2 değerleriyle kategorize edilmiştir.

Sonuç olarak 1 ve 2 değerlerinin oluşturduğu sınıflı bir veri kümesi elde edilmiş oldu.

3.1. Weka Programı ile Elde edilen Bulgular: Karar Ağacı C4.5 Algoritması

KA yönteminin çözümlenmesinde, makine öğrenmesi çalışmalarında sıklıkla başvurulan WEKA programında, *Quinlan* tarafından geliştirilen C4.5 algoritmasının WEKA versiyonu olan J48 yöntemi kullanılmıştır. WEKA, Yeni Zelanda Waikato üniversitesinde geliştirilen bir VM çözüm programıdır. 'The WEKA Explorer' ile kullanıcı kolaylığına sahip bir programdır. Bu program kullanıcıların aktif olarak bir KA oluşturmalarını sağlar. Bir veri alanıyla birlikte istediğinizi seçebileceğiniz seçilmiş iki değişken sunar. Sınıfları çok iyi ayrılmış bir çift değişken bulunduğu verilerin uygun veriler etrafında bir poligon çizerek 2'li bir bölünme yapma olanağı sağlar.²⁷ Aşağıda WEKA'daki J48 algoritmasından elde edilen sonuçlar yer almaktadır.

```

J48 pruned tree
-----
GSYIH <= 0.2605: KUCUK (4.0)
GSYIH > 0.2605
| ENFLASYON <= -1.491: BUYUK (8.0)
| | ENFLASYON > -1.491
| | | IHRACAT <= 0.7102
| | | | SERMAYE_YAT <= 0.117: BUYUK (4.0)
| | | | SERMAYE_YAT > 0.117: KUCUK (2.0)
| | | | IHRACAT > 0.7102: KUCUK (3.0)
Number of Leaves :      5
Size of the tree :      9
Time taken to build model: 0.03 seconds
=== Evaluation on test split ===
=== Summary ===
Correctly Classified Instances   3      60 %
Incorrectly Classified Instances 2      40 %
Kappa statistic                 0
Mean absolute error             0.42
Root mean squared error        0.5745
Relative absolute error        85.9091 %
Root relative squared error    116.7808 %
Total Number of Instances      5

```

Şekil 1. J48 Algoritması sonuçları

²⁷ Ian H. Witten-Eibe Frank, *Data Mining Practical Machine Learning Tools and Techniques*, Second Edition, Morgan Kaufmann Publishers, San Francisco 2005, s.62-63.

Şekil 1’de görüldüğü gibi modelin sınıflama doğruluğunu tespit etmek için verilerin %80’i eğitim kümesi için % 20’si de sınama (test) için kullanılmıştır.

KA incelendiğinde, ilk ayırım *GSYIH* değişkeni olarak belirlenmiştir ve sonra ikinci seviyede, ayrımlar sırasıyla *Enflasyon oranı* ve *İhracat*’ tır. Ağaç yapısında, bir değişken sütunu, belli bir yaprağa atanmış, o yaprağa ulaşan olayların sayısı ile takip edilmiş, algoritmanın kesirli olayları eksik değerleri, karşılamak için kullandığı algoritma yüzünden ondalık sayı ile ifade edilmiş sınıf etiketini ortaya çıkarır.²⁸ Yani parantezin içindeki ilk sayı veri kümesindeki kaç olayın bu düğüm için doğru olarak sınıflandırıldığını, parantezin içindeki ikinci sayı, düğüm tarafından yanlış olarak sınıflandırılan olayların sayısını göstermektedir.

Örneğin, ‘*GSYIH* \leq 0.2605: KUCUK (4.0)’ ifadesinde yer alan (4.0) değeri yanlış sınıflandırılarak ilgili yaprağa gelen olay sayısının bulunmadığını gösterir. Ancak bu değer örneğin (4.2) olsaydı, 2 tanesi yanlış sınıflandırılmış olarak o yaprağa ulaşan 4 olayın olduğu anlamına gelir.

Ağaçta yer alan yaprak sayısının 5 (Number of Leaves: 5) ağaç boyutunun ise 9 (Size of the tree: 9) olduğunu görmekteyiz. Düzensizlik matrisi (confusion matrix), ‘küçük eşit’ sınıfının iki olayının ‘büyük’ sınıfına atandığını ve ‘büyük ‘ sınıfından ‘küçük eşit’ sınıfına atanan olay sayısının olmadığını göstermektedir.

Değerlendirme modülü sınıflandırma hatası ile birlikte aynı zamanda Kappa istatistiğini de çıkarır.²⁹ Ortalama mutlak hata (mean absolute error) 0.42, ve ağaç tarafından atanan sınıf olasılık tahminlerinin hatasının karekök ortalaması (root mean squared error) 0.5745 ve önceki olasılıklara bağlı olan göreceli hata (relative errors) %85.90 olarak bulunmuştur. Hatanın karekök ortalaması, ortalama karesel (kuadratik) kaybın ortalamasının kareköküdür. Ortalama mutlak hata, benzeri şekilde karesi alınmış farklılık yerine mutlak değer kullanarak hesaplar.³⁰

Yine yukarıdaki tabloda görülebileceği gibi, verilerin % 80’ni öğrenme için ve %20’si test için seçildiğinde, doğru sınıflandırılan olay sayısı (türevlenen modelin eğitim kümesinde ne kadar iyi çalıştığını) %60 olarak görülmektedir. Verilerin %66’sı öğrenme ve %20’si test için

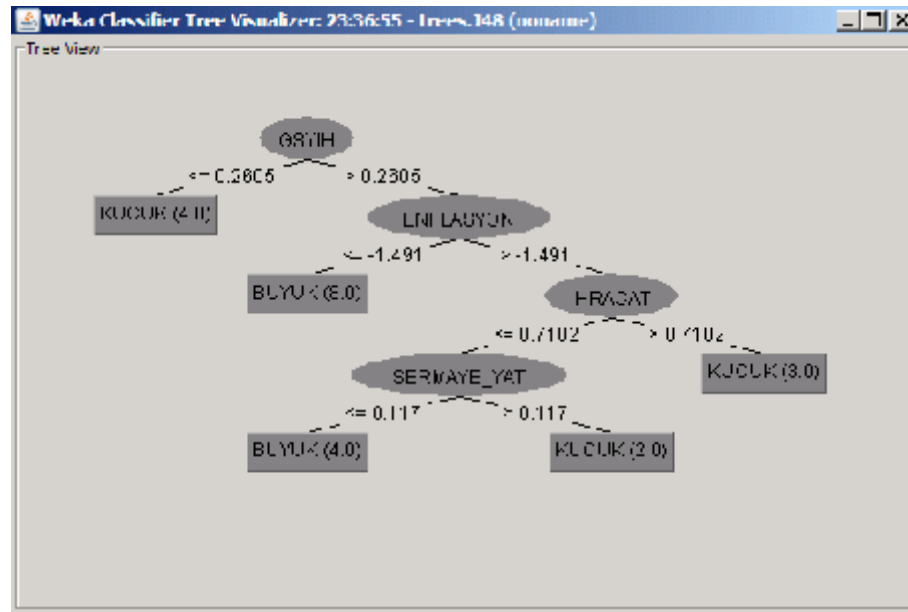
²⁸ Witten-Frank, a.g.e., s. 376.

²⁹Kappa istatistiği, bağımlı gruplarda (önce-sonra karşılaştırmaları) oranların karşılaştırılması için, bir ki-kare testi uyarlaması olan Mc Nemar testinde, tanısal yöntemler arasındaki uyumu test etmek istendiğinde uyumun derecesini ölçer.

³⁰ Witten-Frank, a.g.e., s. 376.

seçildiğinde ise doğru sınıflandırılan olay sayısı %62.50 olurken, ortalama mutlak hata 0.37'ye düşmektedir. Her iki sınıflandırmadan elde edilen doğruluk değeri daha önceki KA uygulamalarında elde edilenlere göre düşüktür.

KA kurallarının daha anlaşılır olması için kurallar bu şekil üzerinde anlatılacaktır. Karar kuralları Şekil 2'deki gibidir.



Şekil 2. J48 Sonuçlarının WEKA programında görselleştirilmesi

Karar ağacı düğümlerinden elde edilen kurallar:

Kural 1. EGER $GSYIH \leq 0.2605$ İSE

SINIF : 9 yıldan küçük ya da eşit

Kural 2. EGER $GSYIH > 0.2605$ İSE VE $ENFLASYON \leq -1.491$ İSE

SINIF: 9 yıldan büyük

Kural 3. EGER $GSYIH > 0.2605$ İSE VE $ENFLASYON > -1.491$ İSE VE $IHRACAT > 0.7102$ İSE SINIF: 9 yıldan küçük ya da eşit

Kural 4. EGER $GSYIH > 0.2605$ İSE VE $ENF. > -1.491$ İSE VE $IHRACAT \leq 0.7102$ İSE VE $SER_YAT. \leq 0.117$ İSE

SINIF: 9 yıldan büyük

Kural 5. EGER $GSYIH > 0.2605$ ISE VE ENFLASYON > -1.491
ISE VE IHRACAT ≤ 0.7102 ISE VE SER_YAT. > 0.117 ISE
SINIF: 9 yıldan büyük

Yukarıdaki kurallar incelendiğinde ilginç sonuçlara ulaşılmaktadır.

Kural 1'e göre AB'ye aday bir ülkenin *GSYIH*'si 0.2605'ten küçük ya da bu değere eşitse müzakere süreci 9 yıldan kısa sürecektir.

Kural 2'ye göre AB'ye aday bir ülkenin *GSYIH*'si 0.2605'ten büyük ise müzakere sürecinin süresi ile ilgili bilgi verilememektedir. Bu nedenle bilgi kazancı *GSYIH*'ye göre daha düşük, diğer değişkenlere göre daha büyük olan *Enflasyon* değişkenine bakılması gerekmektedir. Burada *GSYIH* 0.2605'ten büyük ve *Enflasyon oranı* -1.491'den küçük ya da eşit olan ülkelerin müzakere sürecinin 9 yıldan daha fazla olacağı gösterilmektedir.

Sınıflama doğruluklarında görüleceği gibi 5 değişken analiz dışı kalmıştır. Algoritmalar anlamsız değişkenleri otomatik olarak dışlayarak, yeni öğrenme sürecinde değişken seçimini kendisi yapmaktadır. Bu AB iç dinamiklerinde bazı makro ekonomik değişkenlerin 'adaylık süresini azaltma ya da artırma da' etkinlik gösteremediği şeklinde yorumlanabilir. Söz konusu 5 değişkenin analiz dışı kalmasının sebeplerini istatistiksel analizler yardımı ile analiz edip, bu yorumun doğruluğunu tespit etmek gerekmektedir.

Burada değişkenlerin analiz dışı kalmasının akla gelen ilk nedeni bağımlı değişken (sınıf değişkeni) ile bağımsız değişkenler arasındaki korelasyonun düşük olması ve belli korelasyon katsayısının altında kalan bağımsız değişkenlerin analiz dışı bırakılmasıdır. Korelasyon katsayısı, değişkenler arasındaki ilişkinin yönünü ve derecesini gösterir. Korelasyon analizinde hangi tekniğin kullanılması gerektiğini bulmak için öncelikle değişkenlerin parametrik olup olmadıkları test edilmiştir. SPSS 16.0 programında yapılan Kolmogorov-Smirnov testinin sonuçlarına göre bazı değişkenlerin normal dağılmadıkları görüldüğünden parametrik olmayan yöntemlerden Spearman Sıra Korelasyon Katsayısı hesaplanmıştır. Hipotezin doğruluğunu sınamak için SPSS paket programında Spearman Sıra Korelasyon Katsayısı hesaplanarak, sonuçları Tablo 1'de verilmiştir.

Tablo 1. Spearman sıra korelasyon testi sonuçları

Spearman Sıra Korelasyon Katsayısı	Sınıf Değişkeni (Yıl Farkı)
Enflasyon Oranı	-,239
İhracat	-,125
İthalat	,071
Kur Oranı	,097
Sabit Sermaye Yatırımları	,050
İşsizlik Oranı	-,228
İşgücü	,027
GSYİH	,241
Nüfus Yoğunluğu	-,062
Sınıf değişkeni (Yıl farkı)	1,000

Tablo 1’de görüldüğü gibi WEKA programında analize alınan en önemli 3 değişken Spearman korelasyon analizinde katsayıları en yüksek çıkan değişkenlerdir. *GSYİH* 0.241, *Enflasyon* 0.239 (negatif yönlü), *İhracat* 0.125 (negatif yönlü) olmak üzere en yüksek korelasyon katsayısını veren değişkenlerdir. Şekil 2’deki WEKA programı ile elde edilen KA’da da görüldüğü gibi katsayıların büyüklüklerinin korelasyon katsayılarının sıralaması ile aynıdır.

3.2. Weka Programı ile Elde edilen Bulgular: Yapay Sinir Ağı Çok Katmanlı Algılayıcı Yöntemi

Şekil 3. Çok katmanlı algılayıcı sonuçları

Run information	Devamı
Scheme: weka.classifiers.functions.MultilayerPerceptron -L	Attrib ENFLASYON -2.515475412748431
0.3 -M 0.2 -N 500 -V 0 -S 0 -E 20 -H a	Attrib IHRACAT -1.2781383139706293
Relation: noname	Attrib İTHALAT 1.49490991218534
Instances: 21	Attrib KUR 1.0782245693702304
Attributes: 10	Attrib SERMAYE_YAT 0.6398151456170815
ENFLASYON	Attrib KUR 1.1730475674887562
IHRACAT	Attrib SERMAYE_YAT 0.5164874069702479
İTHALAT	Attrib İSSİZLİK 0.21623578953227346
KUR	Attrib İSGUCU -0.5822147693182445
SERMAYE_YAT	Attrib GSYİH -0.11014247220103966
İSSİZLİK	Attrib NUFUS 1.4193410130382391
İSGUCU	Sigmoid Node 5
GSYİH	Inputs Weights
NUFUS	Threshold 0.13977555166985195
SİNİF	Attrib İSSİZLİK 0.3681256530235419
Test mode: split 80.0% train, remainder test	Attrib İSGUCU -0.7754696086372382
==== Classifier model (full training set) ====	Attrib GSYİH -0.13320531819063422
Sigmoid Node 0	Attrib NUFUS 1.6153358422406137
Inputs Weights	Sigmoid Node 6
Threshold -4.7796732930454375	Inputs Weights
Node 2 3.551331443350244	Threshold 1.816869532208303
Node 3 0.34799920439536297	Attrib ENFLASYON -9.394608649429198
Node 4 1.523171131990317	Attrib IHRACAT -3.8422897825508473
Node 5 1.8217031278229934	Attrib İTHALAT 3.2727687594024952

Node 6	7.437091038667218	Attrib KUR	1.3942388294345023
Sigmoid Node 1		Attrib SERMAYE_YAT	2.49704318777389
Inputs Weights		Attrib ISSIZLIK	1.984875510396493
Threshold	4.768399309849337	Attrib ISGUCU	-3.7290751029086957
Node 2	-3.5074464527199707	Attrib GSYIH	-0.17983621066112435
Node 3	-0.36548772763707055	Attrib NUFUS	5.011785830224099
Node 4	-1.4995026072093676	Class BUYUK	
Node 5	-1.7363670308242918	Input	
Node 6	-7.486915679501787	Node 0	
Sigmoid Node 2		Class KUCUK	
Inputs Weights		Input	
Threshold	0.7363906645866063	Node 1	
Attrib ENFLASYON	-5.100919328998313	Time taken to build model:	0.32 seconds
Attrib IHRACAT	-2.227314727721632	=== Evaluation on test split ===	
Attrib ITHALAT	2.04512415459353	=== Summary ===	
Attrib KUR	1.1065647713447955	Correctly Classified Instances	4 100 %
Attrib SERMAYE_YAT	1.185959152969123	Incorrectly Classified Instances	0 0 %
Attrib ISSIZLIK	0.7464017204347003	Kappa statistic	1
Attrib ISGUCU	-1.7158861343049536	Mean absolute error	0.0299
Attrib GSYIH	-0.1660038325843295	Root mean squared error	0.0409
Attrib NUFUS	2.6512337222030764	Relative absolute error	5.9712 %
Sigmoid Node 3		Root relative squared error	8.0759 %
Inputs Weights		Coverage of cases (0.95 level)	100 %
Threshold	-0.507445559437822	Mean rel. region size (0.95 level)	62.5 %
Attrib ENFLASYON	-0.37885704698366984	Total Number of Instances	4
Attrib IHRACAT	-0.5200093807098264		
Attrib ITHALAT	0.6543708210364358	=== Detailed Accuracy By Class ===	
Attrib KUR	0.8226700053122281	TP Rate	FP Rate
Attrib SERMAYE_YAT	0.16518979320729335	Precision	Recall
Attrib ISSIZLIK	0.07987512252198516	F-Measure	ROC
Attrib ISGUCU	-0.12230345840058515	Area	Class1
Attrib GSYIH	0.09709355999444641	1	BUYUK
Attrib NUFUS	0.8725261191887057	1	0
Sigmoid Node 4		KUCUK	1
Inputs Weights		Weighted Avg.	1 0 1 1 1 1
Threshold	-0.07080578175310887		
Attrib ENFLASYON	-2.1711055194164994		
Attrib IHRACAT	-1.081880369605599		
Attrib ITHALAT	1.331751127274674		

Şekil 3. Çok katmanlı algılayıcı sonuçları

ÇKA, Geri yayılım algoritmasını kullandığı için, çıktı düğümlerinde hata oranı belirlenmiş daha sonra düğümlerin ağırlıkları değiştirilerek en düşük hata değerini veren yapıda durulmuştur. Şekil 3'te görüldüğü gibi C4.5 algoritmasının çözümüyle %60 olan sınıflama başarısı, ÇKA yöntemiyle yine verilerin %80'i eğitim %20'si test seti olarak kullanıldığında %100'e ulaşmıştır. Bu, sınıflama yöntemlerinde nadir ulaşılabilecek bir sınıflandırma başarısıdır.

Şekil 3'te görüldüğü gibi transfer fonksiyonu olarak Sigmoid fonksiyonu kullanılmıştır. Girdi değerleri gizli katmandaki değerlere bağlanmıştır. Her düğüme modelin belirlediği bir ağırlık atanmaktadır. Düğümler kendilerine gelen girdilerin ağırlıklarının toplamından

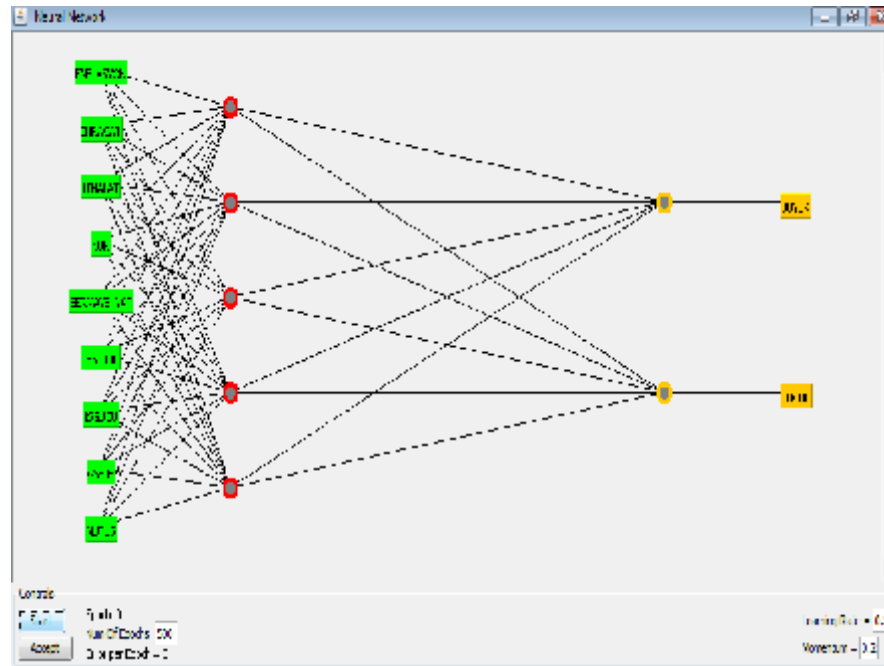
oluşmaktadır. Örneğin, 5 numaralı düğüme gelen girdiler ve ağırlıkları Tablo 2’de gösterilmiştir.

Tablo 2. 5 Numaralı düğüme gelen girdiler ve ağırlıkları

Girdi	Ağırlığı
İŞSİZLİK ORANI	0.3681256530235419
İŞGÜCÜ	-0.7754696086372382
GSYİH	-0.13320531819063422
NÜFUS YOĞUNLUĞU	1.6153358422406137

Ortalama mutlak hata (mean absolute error) 0.029, ve sınıf olasılık tahminlerinin hatasının karekök ortalaması (root mean squared error) 0.040 ve önceki olasılıklara bağlı olan göreceli hata (relative errors) %5.97 olarak bulunmuştur. Hata oranlarının çok düşük olması da sınıflandırma başarısını göstermektedir.

Weka programının ÇKA yönteminde sunmuş olduğu değiştirge ayarları, sonuçları görsel olarak analiz etmemize de olanak tanır. Grafikselle Kullanıcı Ara yüzü (GUI) ile elde edilen görüntü Şekil 4’te verilmiştir.



Şekil 4. ÇKA Sinir ağı görünümü

Şekil 4'te gizli katman ile birlikte grafiksel kullanıcı ara yüzü gösterilmektedir.

4. SONUÇ

YSA ve KA yöntemleriyle elde edilen bulgular karşılaştırıldığında, sınıflandırma başarısı açısından YSA uygulamasının KA uygulamasına üstünlüğü açıkça görülmektedir. YSA'nın sınıflama doğruluğu %100 iken, KA yönteminin uygulama doğruluğu yaklaşık %60'dır. Klasik doğrusal modelleme anlayışından farklı olarak tüm etkenleri ele alan ve veri eğitimleri sonrasında tahmin işleminde analiz ve sınıflandırmayı etkin bir şekilde yapan YSA ideal bir model olarak elde edilmiştir.

YSA başarısındaki etkenlerin başında verileri kullanarak eğitim süreci gerçekleştirmesi ve bu süreç sonrasında kendini daima yenileyerek hata oranını azaltılmaya çalışması gelebilir.

Ancak YSA'da ağıın davranışı açıklanamamaktadır. Bir probleme üretilen çözümün nasıl gerçekleştiğine dair bir bilgi bulunmamaktadır. Ayrıca YSA, KA gibi karar kuralları oluşturamaz. Karar kurallarının oluşturulması kullanıcıya hem pratik bilgi hem de görsel analiz avantajı sağlar.

Çalışma, uygulama amacı açısından değerlendirildiğinde, AB üyeliğinde ülkelerin ekonomik gelişmişlik düzeyinin etkisinin sınırlı kaldığı görülmektedir. Karar kurallarından yola çıkarak en etkili değişkenlerin GSYİH, Enflasyon, İhracat ve Sermaye Yatırımları olduğu gözlenmiştir. Bazı AB ülkelerinin 2011 yılında yaşadığı ve halen devam etmekte olan ekonomik darboğaz düşünüldüğünde, önümüzdeki yıllarda AB üyeliği için ekonomik gelişmişlik kıstasının daha fazla önem kazanacağı düşünülmektedir.

KAYNAKÇA

Bishop, C. M., *Neural Networks for Pattern Recognition*, Oxford University Press, London 1995.

Brada, J.C. and A.M. Kutan, 'The Convergence of Monetary Policy between Candidate Countries and the European Union', *Economic Systems*, Vol. 25, No. 3, 2001.

Bruss, Fritz, 'Benefits and Dangers of EU Enlargement', *Empirica*, Kluwer Academic Publishers, Vol 29, Netherlands 2002.

Doğan, Nurhan ve Kazım Özdamar, 'Chaid Analizi ve Aile Planlaması ile Bir Uygulama', *T. Klin Tıp Bilimleri Dergisi*, Cilt. 23, Eskişehir 2003.

Dormen, Doruk, Bankacılık Sektöründe Müşteri İlişki Yöntemi: CRM Açısından Veri Madenciliği Yöntemi, Doktora tezi, İstanbul 2003.

Doyle, Orla and Jan Fidrmuc, 'Who is in favor of Enlargement? Determinants of Support for EU Membership in the candidate countries' referanda', <http://www.fidrmuc.net/research/refd.pdf>, (09.10.2011).

Dunham, M. H., *Data Mining Introductory and Advanced Topics*, Pearson Education Inc., Southern Methodist 2003.

German, G. W. H., G. and West-M. Gahegang, Statistical and AI Techniques in GIS Classification: A Comparison, http://divcom.otago.ac.nz/Sirc/conferences/1999/14_German.pdf, (11.10.2011).

Halkos, G. and E. Tzeremes, 'Economic Efficiency and Growth in the EU Enlargement', *Journal of Policy Modeling*, Vol. 31, 2009.

Haykin, Simon, *Neural Networks and Learning Machines*, Third edition, Pearson Education, 2009.

Kandartzic, Mehmed, *Data Mining Concepts Models and Algorithms*, John Wiley&Sons, 2003.

Kumar, P., Nitin-V., Sehgal and D.S. Chauhan, 'Performans Evaluation of Decision Tree versus Artificial Neural Network Based Classifiers in Diversity of Datasets', *World Congress on Information and Communication Technologies(WICT 2011)*, Mumbai 2011.

Öztemel, Ercan, *Yapay Sinir Ağları*, Papatya Yayıncılık, İstanbul 2006.

Quinlan, J.R, *C4.5 Programs For Machine Learning*, Morgan Kaufman Publisher, California 1993.

Seidman, Claude, *Data Mining with Microsoft SQL Server 2000*, Microsoft Press, Washington 2001.

Tan, P.N., Steinbach, M. and V. Kumar, *Introduction to Data Mining*, Pearson Addison Wesley, Boston 2006.

Tireli, Tina and Daniela Pessani, 'Use of Decision Tree and Artificial Neural Network Approaches to Model Presence/Absence of *Telestes Muticellus* in Piedmont', *Rever Research and Applications*, Vol. 25, Torino 2009.

Witten, Ian H. and Eibe Frank, *Data Mining Practical Machine Learning Tools and Techniques*, Second Edition, Morgan Kaufmann Publishers, San Francisco 2005.