

BİLGİ EDİNME HAKKI YASASI ÇERÇEVESİNDE YAPILAN ELEKTRONİK BAŞVURULARIN YAPAY SİNİR AĞLARI İLE SINIFLANDIRMASI

Yavuz KILAĞIZ^(*)
Ahmet BARAN^(**)

Özet: Bu çalışmada, Bilgi Edinme Hakkı Yasası çerçevesinde örgütlere yapılan elektronik başvuruların sınıflandırılması esnasında ortaya çıkan işgücü ve zaman kaybı gibi problemleri gidermek amacıyla bir elektronik başvuru sınıflandırma sistemi geliştirilmiştir. Sistem, Erzincan Üniversitesine yapılan elektronik başvuruları sınıflara ayırıp ilgili Daire Başkanlıklarına iletmektedir. İş akışını önemli ölçüde hızlandırıp, iş gücü kaybının önüne geçen Yapay Sinir Ağı destekli sistem, kelime tabanlı olup, danışmanlı öğrenmekte ve %98 doğrulukla çalışmaktadır.

Anahtar Kelimeler: Doküman Sınıflandırma, Yapay Sinir Ağları, Bilgi Yönetimi

Abstract: In this paper, within the framework of the Act on the Right of Information Acquirement, an electronic application classification system has been developed in order to eliminate such problems as the waste of workforce and time during the classification of electronic applications to the institutions. The system first classifies the electronic applications to Erzincan University then forwards them to the relevant Chambers of Affairs. Artificial Neural Network based system, which considerably accelerates the flux of work and hinders the waste of workforce, is a word based and supervised learning system and it works with 98% accuracy.

Keywords: Document Classification, Artificial Neural Networks, Information Management.

I.Giriş

24 Nisan 2004 tarihinde yürürlüğe giren ve kamu kurum ve kuruluşları ile kamu kurumu niteliğindeki meslek kuruluşlarının faaliyetlerinde uygulanan Bilgi Edinme Hakkı Yasası gereğince, kurum ve kuruluşlar, elektronik ortamda veya diğer iletişim araçlarıyla kendilerine yapılan başvuruları değerlendirip, kanunda belirlenen süre içerisinde talep eden gerçek veya tüzel kişiye cevap vermek zorundadır (Resmi Gazete, Kanun No:4892, 2003). Yapılan müracaatların çoğu, kolaylığı nedeniyle elektronik posta yoluyla yapıldığından, kamu kurum ve kuruluşlarının neredeyse tümü internet sitelerinde bilgi edinme hakkı başvurusu yapılabilecek sayfalar hazırlamışlardır. Bu sayfalar aracılığı ile orta ve büyük ölçekli kurum ve kuruluşlara, her gün yüzlerce bilgi edinme başvurusu gelmektedir. Gelen başvuruları sınıflandırarak kurumun ilgili birimlerine yönlendirme işlemi, çoğunlukla görevli personeller tarafından yapılmaktadır. Bu durum, işgücü kaybına neden olduğu gibi, yoğunluğun fazla

^(*)Yrd.Doç.Dr. Erzincan Üniversitesi Sağlık Hizmetleri MYO

^(**)Yrd.Doç.Dr. Erzincan Üniversitesi MYO

olması halinde gecikmelere neden olabilmektedir. Gecikmeler iş akışını yavaşlatmakta hatta bazen yasada belirtilen süreler içinde cevap verilemediği için, kurum yasal yükümlülük altına da girebilmektedir. Olayın diğer bir boyutu da, sınıflandırma işleminin insan aracılığı ile yapılması nedeniyle, az da olsa yasada belirtilen şeffaflık, eşitlik ve tarafsızlık ilkeleri ile bağdaşmayacak olumsuzluklar ve aksaklıklar yaşanabilmektedir.

Aslında gelen başvuruların ilgili birimlere aktarılması işlemi, bir dokümanın önceden belirlenmiş kategorilerden hangi/hangilerine uygun olduğunun belirlendiği bir sınıflandırma işlemidir. Bu konuda yapılan çalışmalarda verilmiş bir dokümanın insan desteğini en aza indirerek otomatik olarak sınıflandırılması için danışmanlı, danışmansız veya yarı-danışmanlı olmak üzere birçok yöntem kullanılmaktadır (Lai ve Lam, 2001: 78-89). Yapay Sinir Ağları (YSA) (Manevitz ve Yousef, 2007: 1466-1481), Sonek Ağaç Sınıflandırma (Suffix Tree Classification) (Zamir ve Etzioni, 1998: 46-54), Gizli Anlam İndeksleme (Latent Semantic Indexing) (Bo vd., 2008: 900-904), K-ortalama (K-means) (Sahami vd., 1997: 845-851), Lineer Sınıflandırma (Linear Classification) (Sandler, 2005: 256-265) Grup Ortalama (Group Average), Ward's Metodu, Ağırlıklandırılmış Ortalama (Weighted Average) (Burgin, 1995: 562-572) yöntemleri günümüzde en sık kullanılan danışmansız sınıflandırma yöntemleridir. Destek Vektör Makinası (Support Vector Machines) (Vapnik, 2000: 148-157), Genetik Algoritma (Genetic Algorithm) (Song, 2009: 9095-9104), Kohonen Tipi Özdüzenleyici Haritalar (Kohonen Type Self Organizing Maps) (İsa, 2009: 9584-9591), Yapay Sinir Ağları (Guerrero-Bote vd., 2003:287-305), Bulanık c-ortalama (Fuzzy c-means) (Benkhalifa vd., 1999: 561-565), Hiyerarşik Bayes (Hierarchical Bayesian) (Iwayama ve Tokunaga, 1995: 1-9), Bayes Ağları (Bayesian Network) (Lam ve Low, 1997 : 2719-2723), Naive Bayes (Chen vd., 2009 : 5432-5435), En Yakın Komşuluk Algoritmaları (Nearest Neighbors) (Han vd., 2001: 53-65), Karar Ağaçları (Decision Trees) (Bing vd., 200: 20-29), Kural Öğrenme (Learning Rule) (Han vd., 2003: 445-446) ve Kavram Madenciliği (Content Mining) (Lui ve Lu, 2002: 599-604) gibi yöntemler en sık kullanılan danışmanlı sınıflandırma yöntemleridir. Bununla birlikte son zamanlarda yarı-danışmanlı sınıflandırma yöntemleri de kullanılmaya başlanmıştır. Bu yöntemlerde sınıflandırmanın belli aşamalarında danışman kullanılırken, belli aşamalarında da makine öğrenmesi kullanılmaktadır (Cardoso-Cachopo ve Oliveira, 2007: 844-851).

Bu çalışmada Bilgi Edinme Hakkı Yasası çerçevesinde Erzincan Üniversitesi'ne yapılan elektronik başvuruları sınıflandırarak ilgili birimlere gönderen bir doküman sınıflandırma sistemi geliştirilmiştir. Geliştirilen sistem çevrimiçi çalışan, kelime tabanlı ve Çok Katmanlı Algılayıcı tipi YSA kullanan danışmanlı bir sınıflandırma sistemidir.

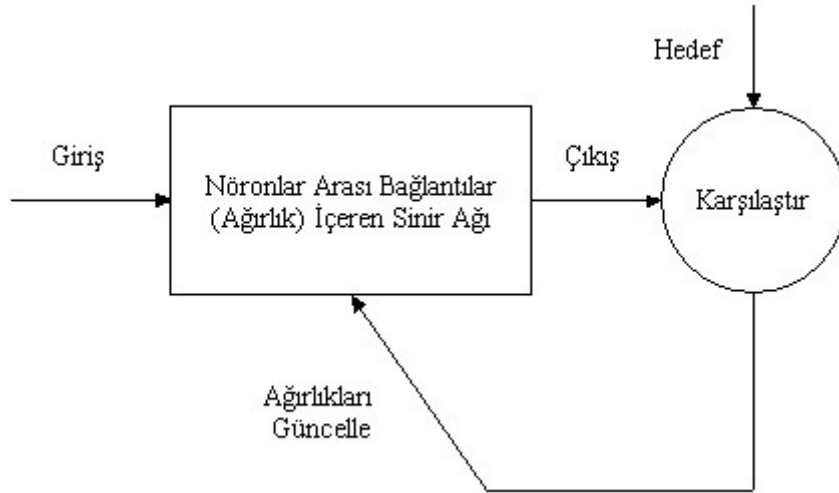
Bu çalışma 4 bölümden oluşmaktadır. Bölüm 2'de yapay sinir ağları tanıtılmış, 3. bölümde ise geliştirilen sistem detaylı bir şekilde anlatılmış ve

simülasyon sonuçları verilmiştir. Elde edilen bulgular sonuç bölümünde tartışılmıştır.

II. Yapay Sinir Ağları

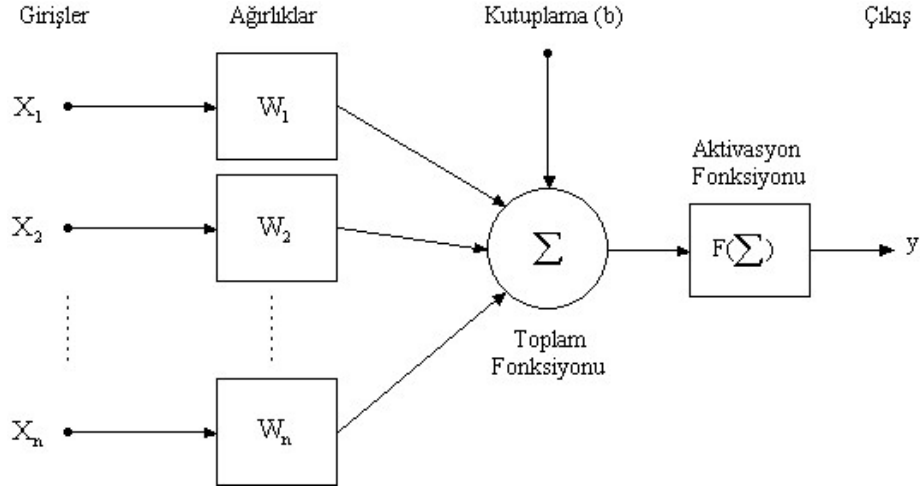
Yapay sinir ağları, insan beyni esas alınarak modellenen sistemlerdir. Başka bir deyişle, beynin yaptığı işlemleri yapabilen, karar veren, sonuç çıkaran, yetersiz veri durumunda var olan mevcut bilgiden yola çıkarak sonuca ulaşan, sürekli veri girişini kabul eden, öğrenen ve hatırlayan algoritmalar (Nabiyev, 2005: 599 -600 ; Haikin, 1994: 47-53 ; Schalkoff, 1997: 23-24).

YSA'lar gerçek hayatta karşılaşılan problemlerde oldukça geniş bir uygulama alanı kazanmışlardır. Bugün, birçok endüstride başarılı şekilde kullanılmaktadırlar. Uygulama alanları için bir sınır yoktur fakat öngörü, modelleme ve sınıflandırma gibi bazı alanlarda ağırlıklı olarak kullanılmaktadır (Sağiroğlu vd., 2003 : 11-15). Şekil 1'de bir YSA'nın temel çalışma prensibi görülmektedir.



Şekil 1: YSA'nın Çalışma Prensibi

YSA, işlem elemanlarının (yapay nöronlar) birbirleri ile çeşitli şekillerde bağlanmasından oluşur ve genellikle katmanlar şeklinde düzenlenir. Bu elemanlar biyolojik sinir sistemlerinden esinlenerek tasarlanmıştır. Belirli bir fonksiyonu gerçekleştirmek için bir sinir ağı, işlem elemanları arasındaki bağlantıların (ağırlıkların) değerleri uygun şekilde ayarlanarak eğitilebilir. Şekil 2'de bir işlem elemanı görülmektedir.



Şekil 2: Bir İşlem Elemanının Yapısı

Her işlem elemanı Şekil 2’de görüldüğü gibi girişler (X_n), ağırlıklar (W_n), kutuplama (b), toplam fonksiyonu (Σ), aktivasyon fonksiyonu $F(\Sigma)$ ve çıkış (y) olmak üzere altı ana kısımdan oluşur (Fu, 1994:79-86; Tsoukalas ve Uhrig, 1997: 125-129).

Girişler, giriş seti veya kendinden önceki bir katmandaki başka bir işlem elemanının çıkışıdır. Burada X giriş bilgilerini içeren vektör olmak üzere $X_1=(X_1, X_2, \dots, X_n)$ şeklinde ifade edilebilir. Ağırlıklar, giriş seti veya kendinden önceki bir katmandaki başka bir işlem elemanının, bu işlem elemanı üzerindeki etkisini ifade eden değerlerdir. Ağırlık katsayı değerlerini içeren vektör ise $W_1=(W_1, W_2, \dots, W_n)$ ile gösterilebilir. Toplam fonksiyonu, girişler ve ağırlıkların tamamının bu işlem elemanına etkisini hesaplayan bir fonksiyondur. İşlem elemanına taşınan her bir bilgi, giriş değerleri ile bağlantı ağırlık değerlerinin çarpılmasıyla elde edilir. Bir işlem elemanında toplanan net bilgi Eşitlik 1 kullanılarak elde edilir.

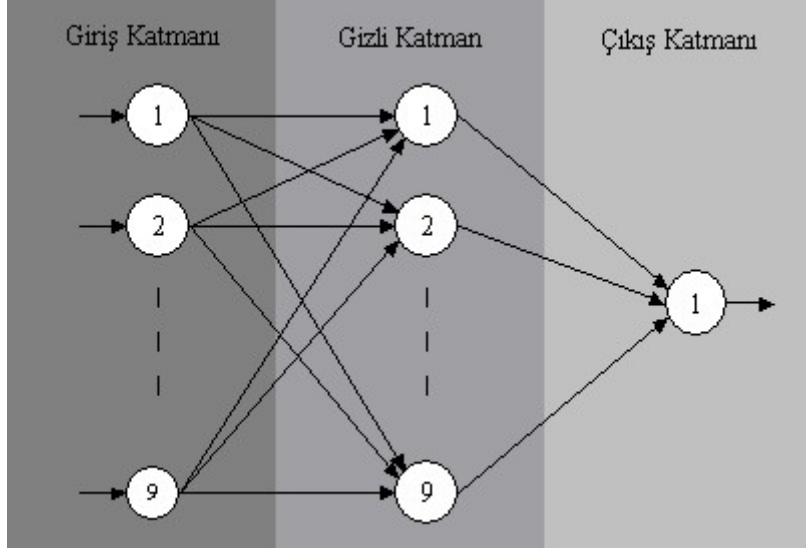
$$NET = \left(\sum_{i=1}^n X_i W_i \right) + b \quad (1)$$

İşlem elemanında toplanan net bilgi bir aktivasyon fonksiyonundan geçirilerek elemanın gerçek çıkış değeri olan y değeri, $y=F(NET)$ ifadesi ile elde edilir. Aktivasyon fonksiyonu, toplam fonksiyonun çıkışında hesaplanan değer 0 ile 1 veya -1 ile 1 arasında nöron çıkış değerine dönüştürüldüğü fonksiyondur (Fu, 1994:146-153). Çıkış, aktivasyon fonksiyonunun sonucudur. Aktivasyon fonksiyonunun seçimi yapay sinir ağının verilerine ve ağın neyi öğrenmesinin istendiğine bağlıdır. Aktivasyon fonksiyonu olarak en çok

sigmoid ve hiperbolik tanjant fonksiyonları kullanılmaktadır. Örneğin eğer ağı bir modelin ortalama davranışını öğrenmesi isteniyorsa sigmoid fonksiyon, ortalama sapmanın öğrenilmesi isteniyorsa hiperbolik tanjant fonksiyonu kullanılması önerilmektedir (Hara ve Nakayama, 1994: 819-822; Kazuyuki ve Kenji, 2000: 41-47). Bu çalışmada ağı, modelin ortalama davranışını öğrenmesi amaçlandığından Eşitlik 2'de çıkış ifadesi verilen sigmoid aktivasyon fonksiyonu kullanılmıştır.

$$y = \frac{1}{1 + e^{-NET}} \quad (2)$$

Ağ topolojisinin belirlenmesi işlemi, yapay sinir ağları kullanımında önemli bir problemdir. Yapay sinir ağlarında, Çok Katlı Algılayıcılar (Multi Layer Perceptron, MLP), Radyal Tabanlı Fonksiyon Ağları (Radial Bases Function Network, RBF), Vektör Kuantalamalı Öğrenme (Learning Vector Quantisation Networks, LVQ), Olasılık Sinir Ağları (Probabilistic Neural Networks, PNN), Genelleştirilmiş Regresyon Ağları (Generalized Regression Neural Network, GRNN), Hopfield Ağı, Elman ve Jordan Ağları, Kohonen Ağı ve Adaptif Rezonans Teorisi (Adaptive Resonance Theory, ART) gibi çok çeşitli ağ yapıları ve modelleri vardır. (Schalkoff, 1997: 142-149). Bu çalışmada sınıflandırma problemleri için en iyi sonuç veren modeller olan MLP ve LVQ modellerinden, MLP modeli kullanılmıştır. (Kohonen, 1990 : 67) MLP modeli, bir giriş, bir veya daha fazla gizli ve bir de çıkış katmanından oluşur (Şekil 3). Bir katmandaki bütün işlem elemanları bir üst katmandaki bütün işlem elemanlarına bağlıdır. Bilgi akışı ileri doğru olup geri besleme yoktur. Bu nedenle ileri beslemeli sinir ağı modeli olarak da adlandırılır. MLP modelinin temel amacı, ağı beklenen çıkışı ile ürettiği çıkış arasındaki hatayı en aza indirmektir. Bu ağlara eğitim sırasında hem girişler hem de o girişlere karşılık üretilmesi gereken (beklenen) çıkışlar verilir (danışmanlı öğrenme). Ağı görevi her giriş için o girişe karşılık gelen çıkışı üretmektir. Veriler, giriş katmanına uygulanır, ara katmanlarda işlenir ve çıkış katmanından da çıkışlar elde edilir. Kullanılan eğitim algoritmasına göre, ağı çıkışı ile arzu edilen çıkış arasındaki hata tekrar geriye doğru yayılarak hata minimuma düşünceye kadar ağı ağırlıkları değiştirilir (Kohonen, 1990: 98-101).

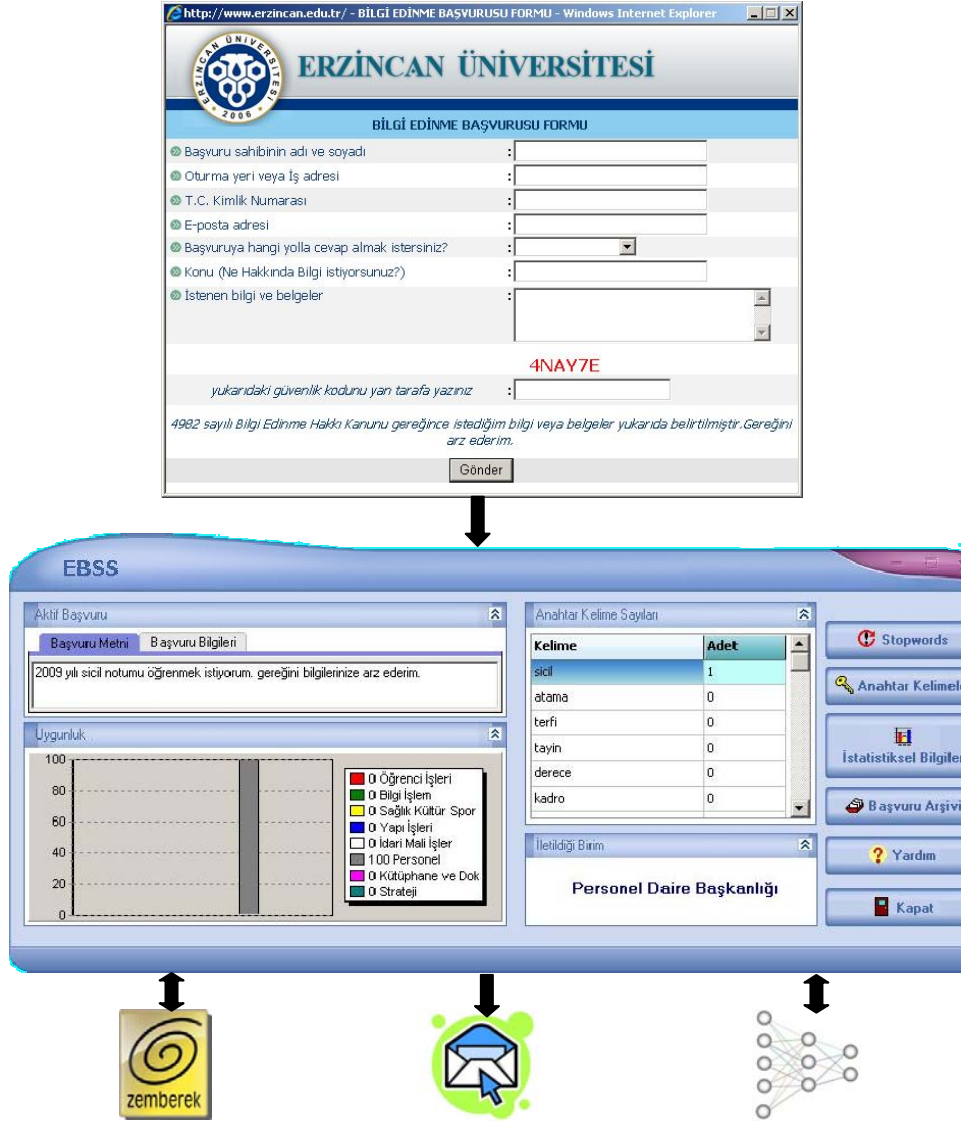


Şekil 3: MLP Modeli

Yapay sinir ağları kullanımı ile ilgili diğer problemler ise, gizli katmanların sayısı, gizli katmanlardaki işlem elemanı sayısı, işlem elemanlarının birbirleriyle bağlantı tipi, öğrenme oranı ve başlangıç ağırlıkları gibi parametrelerin belirlenmeleridir ve bu parametreler oluşturulacak ağın performansını önemli derecede etkilemektedir. Teoride MLP de tek gizli katman kullanılması çoğu problem için yeterli olmaktadır. Diğer parametreler ise çoğunlukla deneme yoluyla bulunur (Cybenko, 1989: 303-309).

III. Sistemin Yapısı ve Çalışması

Çalışmamızda işgücü kaybı, iş akışındaki yavaşlama ve insan kaynaklı kötüye kullanım şeklinde özetlenebilecek olumsuzlukları gidermek üzere Bilgi Edinme Hakkı Yasası çerçevesinde Erzincaan Üniversitesine gelen elektronik başvuruları sınıflandırarak ilgili birimlere ileten bir sistem geliştirilmiştir. Blok şeması Şekil 4’de verilen Elektronik Başvuru Sınıflandırma Sisteminde (EBSS), elektronik başvurular, Erzincaan Üniversitesi Bilgi Edinme Başvuru Formu (BEBF) internet sayfasından EBSS arayüzüne iletilmektedir. Delphi 2009 programlama dili ile hazırlanmış arayüzde, aktif elektronik başvuru metni, BEBF aracılığı ile kullanıcıdan elde edilmiş başvuru bilgileri, anahtar kelimelerin başvuru metninde geçme sayıları ve sınıflandırma sonucu görüntülenmektedir. Ayrıca arayüz aracılığı ile stopword ve anahtar kelime veritabanlarının güncellenmesi yapılabilmekte, geçmiş başvuru bilgileri ve istatistikleri izlenebilmektedir.



Şekil 4: Elektronik Başvuru Sınıflandırma Sistemi

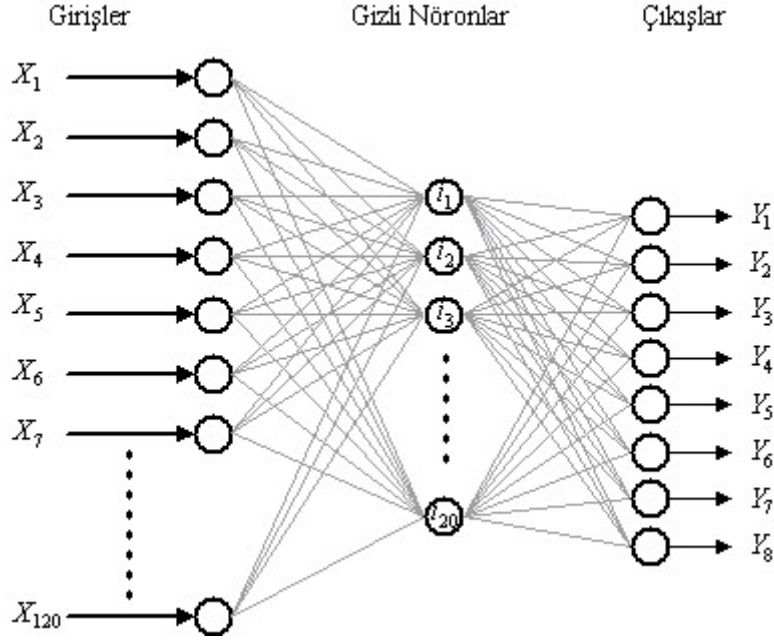
Bu çalışmada, Erzurum Üniversitesi Rektörlüğüne bağlı 8 adet Daire Başkanlığı, sistemimizin 8 sınıfı olarak belirlenmiştir. Literatürde kelime tabanlı doküman sınıflandırma problemlerinde optimal çözüm için seçilecek anahtar kelime sayısının 10-15 arasında olması önerilmektedir (Lewis, 1992: 212-217). Bu nedenle her bir sınıf için ilgili Daire Başkanlığındaki uzmanlarca

belirlenen 15'er adet anahtar kelime tespit edilmiştir. Tablo 1'de Daire Başkanlıkları için örnek bazı anahtar kelimeler verilmiştir.

Tablo 1: *Daire Başkanlıklarına Ait Örnek Anahtar Kelimeler*

Daire Adı	Anahtar Kelimeler
Öğrenci İşleri	transkript
	mezuniyet
	diploma
Bilgi İşlem	bilgisayar
	internet
	otomasyon
Sağlık Kültür Spor	harç
	yemek
	sağlık
Yapı İşleri	ihale
	bina
	inşaat
İdari Mali İşler	ihale
	lojman
	satınalma
Personel	sicil
	atama
	terfi
Kütüphane ve Dokümantasyon	tanıtım
	yayın
	kitap
Strateji	bütçe
	ödeme
	strateji

Oluşturulan YSA, her bir anahtar kelime bir giriş olmak üzere 120 girişe ve her bir sınıf bir çıkış olmak üzere 8 çıkışa sahiptir. Yapılan denemeler sonucunda bu sınıflandırma işlemi için en uygun gizli katman sayısı 1 ve bu gizli katmandaki nöron sayısı 20 olarak tespit edilmiş ve oluşturulan yapay sinir ağı Şekil 5'de verilmiştir.



Şekil 5: Sınıflandırma İşleminin Yapıldığı YSA'nın Yapısı

EBSS arayüzü, gelen her bir elektronik başvurunun giriş özellik vektörünü oluşturmak için önce, elektronik başvuruyu sınıflandırmaya katkısı olmayan ve literatürde stopword (ve, veya, hatta gibi) olarak adlandırılan kelimelerden arındırır. İngilizce dili için kapsamlı bir stopword listesi hazırlanmış olmasına rağmen Türkçe için literatürde en fazla 147 kelimedenden oluşan bir stopword listesi bulunmaktadır (Can vd., 2008: 1-28) ve EBSS bu 147 kelimedenden oluşan stopword veritabanını kullanmaktadır.

EBSS'de stopwordlerden ayıklanan metin, Zemberek yazılımına gönderilerek, kalan kelimelerin kökleri ve ekleri ayrılmaktadır. Zemberek, Türkçe kelime denetleme, kelime çözümleme, kelime önerme, oluşturma, Türkçe karakter kullanılmadan yazılan yazıların dönüştürülmesi ve heceleme gibi çeşitli problemlerin çözülmesi için oluşturulmuş açık kaynak kodlu, platform bağımsız bir yazılımdır (<https://zemberek.dev.java.net/>).

Sistemde daha sonra elde edilen kelime köklerinden elektronik postaya ait giriş özellik vektörü hazırlanmaktadır. YSA'lar ile yapılan sınıflandırma çalışmalarında giriş özellik vektörünün hazırlanmasında İkili (Binary) ve Olasılık (Probabilistic) olmak üzere iki yaklaşım kullanılmaktadır (Özgür vd., 2004: 1819-1831). Giriş özellik vektörü X ,

$$X = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1n} \\ \dots & \dots & \dots & \dots \\ x_{m1} & x_{m2} & \dots & x_{mn} \end{bmatrix}$$

olmak üzere ikili yaklaşımda ağırlıklar Eşitlik 3 aracılığı ile bulunur.

$$x_{ij} = \begin{cases} 1, & \text{eğer } i. \text{ kelime } j. \text{ elektronik başvuruda varsa} \\ 0, & \text{yoksa} \end{cases}, \quad (3)$$

Olasılık yaklaşımında ise metnin uzunluğu (stopwords hariç) ve anahtar kelimenin metinde geçme sayısı göz önünde bulundurularak Eşitlik 4 yardımıyla ağırlıklar belirlenmektedir.

$$x_{ij} = \frac{\text{i. kelimenin j. başvuruda tekrarlanma sayısı}}{\text{j. başvurudaki toplam kelime sayısı}} \quad (4)$$

Çalışmamızda giriş özellik vektörü, olasılık yaklaşımı kullanılarak hazırlanmıştır.

YSA'nın eğitilmesi için veritabanında kayıtlı olan 600 adet elektronik başvurunun 360 tanesi kullanılmıştır. Eğitim için, her bir satır bir anahtar kelimenin ağırlığını ve her bir sütun bir eğitim elektronik başvuruyu temsil etmek üzere 120x360 boyutlu bir vektör giriş özellik vektörü olarak, her bir satırı bir sınıfı ve her bir sütunu bir eğitim elektronik başvuruyu temsil eden 8x360 boyutlu bir vektör de çıkış özellik vektörü olarak kullanılmıştır. Eğitim için kullanılan giriş ve çıkış özellik vektörleri Tablo 2'de verilmiştir. Hangi elektronik başvurunun hangi sınıfa ait olduğunu gösteren çıkış özellik vektörü Daire Başkanlıklarındaki uzmanlarca hazırlanmıştır. Eğitim çıkış özellik vektöründe bir elektronik başvurunun bir sınıfa ait olması 1, ait olmaması 0 ile gösterilmektedir. Çalışmamızda eğitim için, mevcut öğrenme algoritmaları içinde en hızlı sonuç veren algoritmalarından biri olan (Zayani vd., 2008: 758-765) Levenberg-Marquardt öğrenme algoritması (Levenberg, 1944: 164-168; Marquardt, 1963: 431-441) kullanılmış ve $2,6 \cdot 10^{-19}$ ortalama karesel hata (mse) ile sistem öğrenmesini yaklaşık 2 saatte tamamlamıştır.

Tablo 3: Test Sonuçları

Elektronik Başvurunun Gerçek Sınıfı		YSA'nın sınıflandırması	
Sınıf Adı	Başvuru Sayısı	Sınıf Adı	Başvuru Sayısı
Öğrenci İşleri	30	Öğrenci İşleri	30
		Sağlık Kültür Spor	1
Bilgi İşlem	30	Bilgi İşlem	29
		Kütüphane Dok.	1
Sağlık Kültür Spor	30	Sağlık Kültür Spor	29
		Öğrenci işleri	1
Yapı İşleri	30	Yapı İşleri	28
		İdari Mali İşler	2
İdari Mali İşler	30	İdari Mali İşler	29
		Strateji	1
Personel	30	Personel	30
Kütüphane Dok.	30	Kütüphane Dok.	30
Strateji	30	Strateji	30

Tablo 3'den görüldüğü üzere sistem toplam 240 adet başvurunun 5 tanesini yanlış sınıflandırmıştır. Yanlış sınıflandırılan başvurular incelendiğinde başvuru metinlerinde kullanılmış anahtar kelimelerde yazım hataları olduğu tespit edilmiştir.

Yine Tablo 3'den görüldüğü üzere, gerçekte Öğrenci İşleri Daire Başkanlığına ait olan 1 adet elektronik başvuru, sistemin bu başvuru için bulunduğu çıkış vektöründe Öğrenci İşleri Daire Başkanlığı ve Sağlık Kültür Spor Daire Başkanlığı sınıflarına ait değerlerin her ikisi de eşik değerini geçtiği için aynı başvuru, hem Öğrenci İşleri Daire Başkanlığı sınıfına hem de Sağlık Kültür Spor Daire Başkanlığı sınıfına dahil edilerek her iki daire başkanlığına da iletilmiştir. Görünüşte yanlış sınıflandırmaya neden olduğu düşünülebilecek eşik değer kullanımı, aslında sistem için bir güvenlik kontrolü olarak eklenmiştir. Burada bir den fazla sınıfa %95 ve üzerinde uyumlu olan bir başvuru, gerçekte birden fazla sınıfı ilgilendirmesi muhtemel bir başvuru olarak değerlendirilmektedir. Geriye kalan 234 başvuru, sistem tarafından gitmesi gereken daire başkanlıklarına, doğru bir şekilde iletilerek yaklaşık % 98 başarı sağlanmıştır.

IV. Sonuç

Çalışmamızda Bilgi Edinme Hakkı Kanunu çerçevesinde örgütlere yapılan elektronik başvuruların ilgili birimlere iletilmesi esnasında ortaya çıkan insan kaynaklı olumsuzlukları gidermek amacıyla bir elektronik başvuru sınıflandırma sistemi geliştirilmiştir.

Sistemin uygulanması ile yapılan başvurular insan gücüne ihtiyaç olmaksızın ilgili birimlere iletilmekte, böylece iş gücü ve zaman kayıpları minimize edilmektedir. Ayrıca sistemle, insan kaynaklı gecikme ve hata gibi

olumsuzluklardan kaynaklanacak yasal yükümlülükler de asgariye indirilebilecektir.

%98 doğrulukla çalışan sistemde, %2'lik hatanın anahtar kelimelerde yapılan yanlış yazılımdan kaynaklandığı tespit edilmiştir. Kelime tabanlı sınıflandırma sistemlerinin genel bir kısıtı olan bu durum, kullanıcı etkileşimli veya etkileşimsiz kelime önerme algoritmaları ya da kelimelerin metinlere etkisini azaltmaya yönelik algoritmalar kullanımı ile iyileştirilebileceği düşünülmekte ve çalışmanın bu yönde geliştirilmesi hedeflenmektedir.

Kaynaklar

- 4982 Sayılı Bilgi Edinme Hakkı Kanunu, Resmi Gazete, Sayı: 25269, 24 Ekim 2003.
- Benkhalifa, M., Bensaïd, A. ve Mouradi, A. (1999) "Text Categorization Using the Semi-Supervised Fuzzy c-Means Algorithm", *18th International Conference of the North American Fuzzy Information Processing Society*, ss.561 565.
- Bing, L., Yiyuan, X. ve Philip, S.Y. (2000) "Clustering Through Decision Tree Construction", *Ninth International Conference on Information and Knowledge Management*, ss.20 29.
- Burgin, R. (1995) "The Retrieval Effectiveness of Five Clustering Algorithms as a Function of Indexing Exhaustivity", *Journal of the American Society for Information Science*, 46 (8), ss.562 572.
- Can, F., Koçberber, S., Balçık, E., Kaynak, C., Öcalan, Ç. ve Vursavaş, O. (2008) "Information Retrieval on Turkish Texts", *Journal of the American Society for Information Science and Technology*, ss.1 28.
- Cardoso-Cachopo, A. ve Oliveira, A.L. (2007) "Semi-supervised Single-label Text Categorization Using Centroid-based Classifiers", *2007 ACM Symposium on Applied Computing Proceeding*, ss.844 851.
- Chen J., Huang H., Shengfeng, H. ve Qu, Y. (2009) "Feature Selection for Text Classification with Naive Bayes", *Expert Systems with Applications*, 36(3), ss.5432-5435.
- Cybenko, G. (1989) "Approximation by Superposition of a Sigmoidal Function", *Math. Control Signals Systems*, 2, ss.303 309.
- Fu, L.M. (1994), *Neural Networks in Computer Intelligence*, McGraw-Hill, OH.
- Guerrero-Bote, V.P., Lopez-Pujalte, C., Moya-Anegon, F.D. ve Herrero-Solana, V. (2003) "Comparison of Neural Models for Document Clustering", *International Journal of Approximate Reasoning*, 34, ss.287 305.
- Haikin, S. (1994), *Neural Networks. A Comprehensive Foundation*, Prentice-Hall, NJ.
- Han H., Manavoglu E., Giles C.L. ve Zha, H. (2003) "Rule-based Word Clustering for Text Classification", *26th annual international ACM*

- SIGIR Conference on Research and Development in Informaion Retrieval*, ss.445-446.
- Han, E., Karypis, G. ve Kumar, V. (2001) "Text Categorization Using Weight Adjusted k-Nearest Neighbor Classification", *Computer Science*, 2035, ss.53 65.
- Hara, K. ve Nakayama, K. (1994) "Comparison of Activation Functions in Multi Layer Neural Networks for Pattern Classification" *ICAN'94*,ss.819 822.
<https://zemberek.dev.java.net/>
- Isa. D., Kallimani, V.P. ve Lee, L.H. (2009) "Using the Self Organizing Map for Clustering of Text Documents", *Expert Systems with Applications*, 36(5), ss.9584 9591.
- Iwayama, M. ve Tokunaga, T. (1995) "Hierarchical Bayesian Clustering for Automatic Text Classification", *International Joint Conference on Artificial Intelligence*, ss.1 9.
- Kazuyuki, H. ve Kenji, N. (2000) "Selection of Activate Function For Multilayer Neural Networks", *Reports of the Tokyo Metropolitan Technical College*, ss.41 47.
- Kohonen, T. (1990), *Self-organization and Associative Memory*, Springer-Verlag, Heidelberg.
- Lai, K. ve Lam, W. (2001) "Automatic Textual Document Categorization Using Multiple Similarity-Based Models", *Lecture Notes in Computer Science*, 2035, ss.78 89.
- Lam, W. ve Low K. (1997) "Automatic Document Classification Based on Probabilistic Reasoning: Model and Performance Analysis", *IEEE International Conference on Systems, Man and Cybernetics*, ss.2719 2723.
- Levenberg, K. (1944) "A Method for the Solution of Certain Nonlinear Problems in Least Squares", *Quarterly of Applied Mathematics*, 2, ss.164 168.
- Lewis, D.D. (1992) "Feature Selection and Feature Extraction for Text Categorization", *Speech and Natural Language*, ss.212 217.
- Liu, R. Ve Lu, Y. (2002) "Incremental Context Mining for Adaptive Document Classification", *Eighth International Conference on Knowledge Discovery and Data Mining*, ss.599 604.
- Manevitz, L. ve Yousef, M. (2007) "One-Class Document Classification via Neural Networks", *Neurocomputing*, 70, ss.1466 1481.
- Marquardt, D.W. (1963) "An Algorithm for Least-Squares Estimation of Nonlinear Parameters", *Journal of the Society for Industrial and Applied Mathematics*, 11, ss.431 441.
- Nabiyev, V. (2005), *Yapay Zekâ, Genişletilmiş İkinci Baskı*, Seçkin Yayıncılık, Ankara.

- Özgür, L., Güngör, T. ve Gürgeç, F. (2004) "Adaptive Anti-Spam Filtering for Agglutinative Languages : A Special Case for Turkish", *Pattern Recognition Letters*, 25, ss.1819 1831.
- Sağiroğlu, Ş., Beşdok, E. ve Erler, M. (2003), Mühendislikte Yapay Zeka Uygulamaları 1 – Yapay Sinir Ağları, Ufuk Yayınevi, Kayseri.
- Sahami, M., Yusufali, S. ve Baldonado, M. (1997) "Real-time Full-text Clustering of Networked Documents", *National Conference on Artificial Intelligence*, ss.845 851.
- Sandler, M. (2005) "On the Use of Linear Programming for Unsupervised Text Classification", *ACM-SIGKDD Conference on Knowledge Discovery and Data Mining*, ss.256 264.
- Schalkoff, J.R. (1997), Artificial Neural Network, McGraw-Hill, OH.
- Song, W., Li, C. H. ve Park, S.C. (2009) "Genetic Algorithm for Text Clustering Using Ontology and Evaluating the Validity of Various Semantic Similarity Measures", *Expert Systems with Applications*, 36(5), ss.9095 9104.
- Tsoukalas, L.H. ve Uhrig, R.E. (1997), Fuzzy and Neural Approaches in Engineering, John Wiley & Sons Inc.
- Vapnik, V.(2000), The Nature of Statistical Learning Theory, Springer, New York.
- Yu, B., Xu, Z. ve Li, C. (2008) "Latent Semantic Analysis For Text Categorization Using Neural Network", *Knowledge-Based Systems*, 21, ss.900 904.
- Zamir, O. ve Etzioni, O. (1998) "Web Document Clustering:a Feasibility Demonstration", *21st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, ss. 46 54.
- Zayani, R., Bouallegue, R. ve Roviras, D. (2008) "Levenberg-Marquardt Learning Neural Network for Adaptive Pre-0 Distortion for Time-Varying hpa with Memory in OFDM Systems", *16. European Signal Processing Conference*, ss.758 765