İstenmeyen Elektronik Postaların (SPAM)Filtrelenmesi için Bir Uzman Sistem Tasarımı ve Gerçekleştirilmesi

Cahide ÜNAL[1], İsmail ŞAHİN[2*]
[1]MEB, Yenilik ve Eğitim Teknolojileri Genel Müdürlüğü
[2]Gazi Üniversitesi, Teknoloji Fakültesi, Endüstriyel Tasarım Mühendisliği
(Geliş/Received :31.08..2016; Kabul/Accepted : 18.10.2016)

ÖZ

Günümüzde internet teknolojilerinin yaygınlaşmasıyla gelişen elektronik haberleşme bazı sorunları da beraberinde getirmiştir. Elektronik haberleşmenin en önemli sorunlarından biri SPAM olarak isimlendirilen istenmeyen mesajların internette yayılmasıdır. Bu çalışmada, istenmeyen elektronik postaları (SPAM) filtrelemek için bir uzman sistem tasarlanmış ve sonuçları ortaya konulmuştur. Öncelikle, SPAM filtreleme yazılımları ile ilgili bir literatür taraması yapılmıştır. Daha sonra tasarlanan uzman sistem uygulaması adım adım anlatılmıştır. Uygulamada ilk olarak SPAM ve normal e-posta olmak üzere 4601 elektronik postadan oluşan bir veri kümesi hazırlanmıştır. Veri kümesindeki her e-postanın içeriği incelenmiş, metinlerdeki nitelikler belirlenmiştir. Geliştirilen sistem, bu niteliklerden seçim yaparak ve e-postaların IP adreslerine bakarak sözkonusu niteliklerin SPAM olup olmadığına karar vermektedir. Çalışmanın başarısı gerçek zamanlı ugulamalar ile ortaya konmuştur.

Anahtar Kelimeler: E-Posta, SPAM, Uzman Sistem, Yapay Zeka.

Design and Implementation of the Expert System for Filtering of Unwanted Electronic Mails (SPAM)

ABSTRACT

Today, the spread of internet technology in developing electronic communication has brought some problems. One of the most important problems of electronic communication on the Internet is to spread unwanted messages known as SPAM. In this study, an expert system is designed for filtering SPAM and its results are given. Firstly, the literature is reviewed related to SPAM filtering softwares. Later, the expert system designed application is explained step by step. A dataset consisting of 4601 electronic mail was prepared including SPAM and normal e-mail. Content of each e-mail in dataset is analyzed and attributes in texts are detected. Developed system decide whether this attributes are SPAM or not by electing from this attributes and by regarding IP address of e-mails. The success of this study is put forth with real time applications.

Keywords: E-Mail, SPAM, Exper System, Artificial Intellegence.

1. GİRİŞ (INTRODUCTION)

İnternetin en çok kullanılan hizmetlerinden biri elektronik haberleşmedir. Ancak internetin gelişmesi ve yaygınlaşması ile birlikte, elektronik haberleşme bir takım sorunları da beraberinde getirmiştir. Elektronik haberleşmenin en önemli sorunlarından biri SPAM olarak isimlendirilen istenilmeyen mesajların internette yayılmasıdır [1].

SPAM veya istenmeyen e-posta kavramı genellikle, reklam, ürün ve web ilanları, pornograf, kolay para kazanma vb. içerikli e-postalar için kullanılır. İnternette ortalama bir kullanıcı bir günde yaklaşık 10-50 SPAM e-posta alır ve gönderilen e-postaların nerdeyse yarısı (yaklaşık 13 milyar adet) kadar istenmeyen ticari e-posta gönderilir [2].

Günümüzde istenmeyen mesajların engellenmesi için veri madenciliği teknikleri ile önlemler alınmaya çalışılmaktadır. Bu amaçla karar destek makineleri, yapay sinir ağları, genetik algoritma veya karınca kolonisi, karar destek vektör makineleri, bayesian sınıflandırıcı gibi algoritmalar kullanılarak otomatik SPAM filtreme araçları geliştirilmiştir [1].

İstenmeyen e-postalardan kurtulmak için çok emek ve zaman harcanmaktadır. Aynı zamanda disk alanı ve bant genişliği gibi kaynakların da önemli ölçüde kullanıldığı bir gerçektir. Ayrıca, kötü niyetli kişiler veya şirketler, çeşitli aldatmacalar ile yasadışı ürünler ve diğer uygunsuz malzemeleri işlemek için istenmeyen e-postaları fırsat olarak kullanmaktadır [2].

İstenmeyen e-postalara karşı çeşitli metotlar kullanılmıştır. Bunlar ikiye ayrılır: statik metotlar ve dinamik metotlar. Statik metotlar, SPAM mesajları önceden hazırlanmış bir adres listesi yardımıyla saptamaya çalışırlar. Dinamik metotlar ise, e-posta mesajlarının içeriğini inceleyerek SPAM mesajların filtreleme algortimalarını bu içeriğe göre hazırlarlar [3].

Bu çalışmada dinamik metotlara göre istenmeyen e-postaları (SPAM), niteliklerden seçim yaparak ve e-postaların IP adreslerine bakarak tespit eden bir uzman sistem tasarlanmıştır.

**Sorumlu Yazar (Corresponding Author)*
e-posta: isahin@gazi.edu.tr
Digital Object Identifier (DOI) : 10.2339/2017.20.2 267-274

Çalışmanın ikinci bölümünde uzman sistemler, üçüncü bölümünde ise SPAM e-posta konusu hakkında ayrıntılı bilgi verilmiştir. Dördüncü bölümde tasarlanan uzman sistemin yapısı ve uygulanması, son bölümde ise sonuç ve önerilere yer verilmiştir.

2. UZMAN SİSTEMLER (EXPERT SYSTEMS)

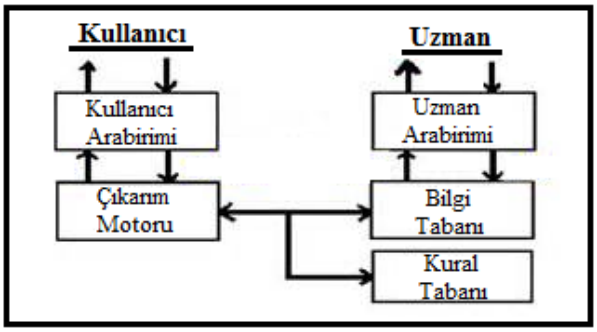
Uzman Sistemler (US), belirli konuda uzman olan bir ve birçok insanın yapabildiği muhakeme ve karar verme işlemlerini modelleyen bir yazılım sistemidir [4].

Günümüzde çok değişik alanlarda uzman sistemler geliştirilmektedir. Bunlar arasında yöneticilere karmaşık planlama problemlerinde yardımcı olan sistemler, hastalık teşhisinde, maden yatakları bulmada ya da karmaşık bilgisayar donanım yapılarını verilen şartlara uygun olarak tasarlamada veya eğitimde kullanılan sistemler olarak sayılabilir [5].

Uzman sistemler ile geleneksel sistemler arasındaki en büyük farklardan birisi muhakeme yeteneğidir. Geleneksel sistemler, uzman sistemlerin aksine muhakeme gerektiren konularda zayıf kalır. Buna karşılık uzman sistemler numerik işlemlerde zayıf kalmaktadır [6,7].

2.1. Uzman Sistemin Genel Yapısı (General Structure of Expert System)

Tasarlanmış bir uzman sistemde algoritma yoktur. Her zaman bilgiye dayalı işlem yapılır. Bilgi tabanından bilgi çağırılır, işlem yapılıp arama gerçekleştikten sonra sonuca varılıp bilgi dâhilinde açıklaması yapılır. Sistem doğru şekilde tasarlanırsa kendini geliştirebilir. Öğrenme yeteneği kazandırılabilir. Bir uzman sistemin genel yapısı Şekil 1'de verilmiştir.



Şekil 1. Uzman sistemin genel yapısı [8] (General Structure of Expert System)

Bilgi Tabanı (Knowledge Base): B: Bilgi mühendisleri tarafından oluşturulur ve bilgi mühendisleri gerçek insan uzmanın bilgisini kurallar ve stratejilere dönüştürür. Bilgi tabanını oluşturan bileşenler kurallar, gerçekler, şebekeler ve çerçevelerden oluşur. Sistem yapısı aynı kalmak koşulu ile konu kapsamı genişletilebilir ve güncellemeler için eklemeler yapılabilir olmalıdır [6].

Kurallar Tabanı (Rule Base): Uzman sistemin bir sonuca varabilmesi için bilgiyi kullandığı, bilgiyi çağırdığı hafızaya "kurallar tabanı" denir. Kurallar genel olarak iki kısımdan oluşmaktadır. Birincisi "Varsayım"; eğer cümlesi ile başlayan ve-veya ile kuralları belirten bir cümle yapısıdır. İkincisi ise "Çıkarım"; O halde kelimesi ile başlayarak kuralların oluşturulduğu kısımdır [8].

Çıkarım Motoru (Inference Engine): Problem verisini örgütler ve uygulanabilir kurallar için bilgi tabanında arama yapar. Çalışan bellek ve çıkarım motoru "Sonuç Çıkarma Mekanizması" olarak adlandırılan ikinci modülü temsil eder. Bu modül çıkarımı gerçekleştirirken hem ileri hem de geriye doğru zincirleme metodu ile sonuca varabilen mekanizmadır [6].

Açıklama Ünitesi (Description Unit): Uzman sistem tasarlamanın ya da bir uzman sistemin en büyük avantajlarından biri, sonuca vardığı şeyin bir sebebi olması ve bunu açıklayabilmesidir [8-10]. Açıklama ünitesi, geliştirilen yazılım aracılığıyla elde edilen sonuçların kullanıcılara iletilmek üzere raporlandığı bölümdür. Raporun doğruluğu geri bildirim için önemlidir.

3. SPAM E-POSTA (SPAM E-MAIL)

SPAM e-posta'ya geçmeden önce e-posta tanımının açıklanması gerekmektedir. E-posta, bireylerin bilgisayar ağına bağlantısını sağlayan ve bir bilgisayardan başka bir bilgisayara internet üzerinden mesaj yollamaya yarayan bir araçtır. E-posta, başlangıçta sadece düz yazı mesajlar göndermek amacıyla geliştirilmişken daha sonra geliştirilen tekniklerle e-posta içinde resim, ses, video, html dokümanları, çalışabilir program vb. kullanımı mümkün hâle gelmiştir [11,12]. e-Posta içerisinde; Gönderen ya da Kimden (From), Alıcı ya da Kime (To), Konu (Subject), Tarih (Date), İleti gövdesi (Body), Karbon Kopya "KK" (CC, Carbon Copy), Gizli Karbon Kopya "GKK" (BCC, Blind Carbon Copy), Yanıtla (Reply), Herkesi yanıtla (Reply All), Yönlendirme (Forward, Fwd), Ek (Attachment) bileşenleri bulunmkatadır [13].

SPAM e-posta ise, çoğunlukla çok fazla sayıda alıcıya gönderilen, talep edilmemiş elektronik iletileri tanımlamak için kullanılan bir terimdir. İlk SPAM, 1 Mayıs 1978 tarihinde DEC'in, ABD'nın Batı kıyısındaki tüm ARPANET adreslerine yeni ürünlerini tanıtmak için gönderdiği e-posta olarak kabul edilmektedir. İlk ciddi SPAM girişimi ise 1994 yılında iki avukatın kendi hizmetlerini anlatan bir reklam iletilerini USENET'teki 1000'lerce gruba göndermeleri olarak kabul edilir.

SPAM e-postaların içerikleri her zaman olmamakla beraber genellikle, ürün ya da hizmet pazarlamak gibi ticari amaçlara hizmet eder [14].

Genel olarak aşağıdaki karakteristiklere sahip olurlar:

- Çoğunlukla alıcıya hiç bir şey ifade etmezler.
- Çirkin ya da yasadışı içerikle gelirler ya da onlara yönlendirirler.
- İçerikleri yalan ya da yanıltıcı olur.
- Mesajın başlık bilgileri tahrip edilmiş olur.
- Alıcıların bu dağıtımdan ileti almak istemediklerini belirtebilecekleri geçerli/ fonk-siyonel bir adres sunmazlar.

- Elde edilmesi ve kullanılması kişilik haklarına tecavüz niteliği taşıyan içeriklere sahip olurlar ya da bu yolla toplanan bilgiyi, kitleyi kategorize etmek için kullanırlar [14].

3.1. Sıkça Rastlanılan Bazı SPAM Mesaj Örnekleri

(Some Common Spam Message Examples)

- Ürün reklamları
- Hemen para kazandıracağını söyleyen SPAM'ler
- Acele kan aranıyor mesajları
- Lösemili hastaya yardım isteyen SPAM'ler.
- Porno içerikli SPAM'ler
- Parti propagandaları
- Satılık binlerce e-posta adresi

3.2. SPAM E-posta Zararları (Spam E-mail Demages)

SPAM türündeki iletilerin yarattığı sorunlar şöyle özetlenebilir:

Gizlilik (Privacy): Hangi eposta adresleri ve hangi kişisel bilgilerin elde edildiği ve biriktirildiği noktasında önemli gizlilik ihlalleri söz konusudur.

İçerik (Content): SPAM e-postaların ciddi bir çoğunluğunun içerdiği yasalara aykırı içerik, pornografi, çocuk pornografisi, yasadışı online kumar servisleri, piramit satışlar, hemen zengin olma vaadleri ile aldatıcı ticari eylemler bireylerin psikolojilerini olumsuz yönde etkilemektedir. Özellikle toplumun küçük yaştaki mensuplarının Internet'in olumlu özelliklerinden ziyade zararlı yönleri ile tanışmasına neden olmaktadır.

Aldatıcı Eylemler, Spoofing (Deceptive Actions, Spoofing): Spoofing, eposta başlıklarının değiştirilmesi ile iletinin orijinal göndericisi yerine başka bir yerden ya da kurumdan geliyormuş gibi gösterme işlemidir. SPAM yapanlar iletilerinin dikkate alınıp okunması ve cevap verilmesi için spoof yaparak ciddi ve saygın kuruluşların isimlerini kullanabilirler. Bu da kurban seçilen kuruluşların ticari ünlerine zarar getirmekte, zaman ve müşteri kayıplarına yol açmaktadır ve bu durumun düzeltilmesi için yapılan harcamalar kuruma ciddi bir mali yük oluşturmaktadır.

Finansal Tutar (Financial Amount): 2001 yılında Avrupa Birliği'nin yaptığı bir araştırmaya göre tüm dünyada bir yılda meydana gelen SPAM faaliyetlerinin tüm internet kullanıcılarına maliyeti olarak 10 milyar dolar dolaylarında [14].

3.3. SPAM Kaynaklı Saldırı Metotları (SPAM-Originating Attack Methods)

SPAM kaynaklı saldırılar çoğunlukla merak uyandırma, reklam, korku, eğlence, politik, yardım gibi konularla karşımıza çıkmaktadır. Saldırı amaçlı kişiler ise hedef olarak öncelikle sistemdeki en kolay giriş yolunu yani açıklıkları denerler.

SPAM postalarla birlikte taşınan virüs vb. yazılımların tespit edilmesi oldukça güçtür. Çok değişik yöntem ve senaryolarla kişinin bilgisayarına gizli bir ajan yazılım olarak yerleşebilmektedir. Kendilerini faydalı bir program olarak göstererek kullanıcının onayını aldığından dolayı çoğu güvenlik önlemleri yetersiz kalabilmektedir.

Uygulamaların sanal ortama taşındığı e-devlet, e-kurum, e-bankacılık gibi işlemlerde kullanılan kullanıcı kimlik doğrulama, hesap ve şifre bilgilerinin elde edilmesine yönelik phishing türü yanlış yönlendirmelerde de kullanılabilmektedir. Gizli DNS ve hosting yanıltma yöntemleri kullanılarak, gelen bir e-postanın gerçek bir bankadan veya kurumdan geldiği izlenimi vermek suretiyle bir takım kişisel bilgileri sahte formlarla istemektedirler. Çok sık yaşanan bu tür mağduriyetlerden dolayı kurumlar web sayfalarından veya SMS ile ilgili, iletinin SPAM olduğuna dair sürekli uyarıda bulunmaktadırlar.

SPAM yoluyla yerleşen zararlı bir yazılım kişinin Outlook'undan adres defterindeki tüm kullanıcılara kendi adında tuzak olabilecek yanıltıcı postalar gönderebilmektedir. Karşıdaki kişi gelen zararlı eklentiye sahip e-postanın tanıdığı ve güvendiği kişiden geldiği varsayımıyla rahatlıkla onay verebilmektedir. Bu şekilde bilgisayarlara yüklenebilen bir takım yazılımlar klavyeden girilen bilgileri, fare ve ekran görüntülerini saldırgan hedefe doğrudan göndermektedir. Zehirlenmiş PC olarak tanımlanan bu bilgisayarlar üzerinden başkalarına ait binlerce reklam maili gönderilebilir. Uzak masaüstü servisini başlatarak saldırganın doğrudan erişmesini sağlar [15].

3.4. SPAM Önleme Teknikleri (SPAM Prevention Techniques)

SPAM e-postalarla mücadelede sunucu ve istemci tarafında alınması gereken bir takım önlemler vardır. İnternet tabanlı veya Outlook benzeri e-posta okuyucularda önemsiz posta klasörü konfigüre edilerek mutlaka tanımlanmalıdır. Doğrudan tanıma ve öğrenme yeteneği sayesinde daha önce bir kez SPAM olarak belirtilen bir posta türü artık SPAM klasörüne gidecektir. Günümüzde Antivirüs yazılımları sadece virüs temizlemekle kalmayıp diğer anti-SPAM, antitrojan, anti-spyware, internet security gibi birçok anti-tarama yeteneği olan tümleşik yazılımlar haline gelmiştir. Kullanılmakta olan güncel antivirus yazılımının SPAM tarama özelliği aktif tutulmalıdır.

Posta hizmeti veren sunucu sistemlerinde ise domain üzerinden gelen giden tüm e-postalar; içerik, konu, kimden ve ekler olmak üzere birçok kritere göre taratılır. Bu amaçla GFI benzeri pek çok ticari SPAM önleme yazılımları geliştirilmiştir.

3.5. SPAM Filtreleme Yazılımları (SPAM Filtering Softwares)

Günümüze SPAM filtreleme yazılımları iki kola ayrılmaktadır. Bu yöntemlerden birincisi önceleri SPAM içerikli e-posta yollayan adresleri filtrelemekle başlayıp günümüzün sıkça kullanılan RBL sistemlerine uzanan teker teker engelleme metodudur [16].

İlk metoda göre SPAM dağıtımına aracılık eden IP adreslerinin engellenmesi yoluyla SPAM gönderiminin önüne

geçilmeye çalışılmaktadır. Bu metodun avantajı gönderilen mesajın içeriğinden bağımsız olarak doğrudan mesajın kaynaklandığı noktaya göre filtreleme yapması sebebiyle içerik üzerinde yapılacak düzenlemeler yoluyla atlatılmasının mümkün olmamasıdır. Buna karşılık SPAM gönderiminde kullanılan sistemleri tespit etmenin zorluğu ve SPAM göndericilerinin çeşitli anonimleştirici sistemler üzerinden bağlantılarını gerçekleştirmeleri sebebiyle engelleme listelerinin yeterince hızlı bir biçimde güncellenmesi mümkün olmamaktadır.

İkinci yöntem olan içerik filtreleme yöntemi ilk olarak e-postaların belirli anahtar kelimeleri içerip içermemesine göre filtrelenmesine dayanır. Kelimelerin teker teker filtrelenmesinin gelen mesajların çokluğu ve çeşitliliği karşısında yetersiz kalmasının yanında bu sistemin normal e-postaları da SPAM olarak tanımlayarak yaptığı hatalı tespitlerin çokluğu sebebiyle mesajları statik kriterler yerine istatistiksel yöntemler kullanarak filtreleyen sistemler geliştirilmiştir. İstatistiksel içerik filtreleme sistemleri, e-posta mesajlarını tek bir kelime yerine içindeki kelimeler arasında tespit edilen istatistiksel bağlantılar yardımıyla sınıflandırır. Günümüzde modern SPAM filtreleme sistemleri bu yöntemleri teker teker uygulamak yerine bu sistemleri birlikte uyarlamakta ve e-posta hakkında nihai kararlarını vermeden önce bu sistemlerden gelen sonuçları değerlendirerek bir karar vermektedirler. Bu sayede sistemlerin tek başlarına ortaya çıkan eksikliklerini gidermek mümkün olmaktadır.

3.6. SPAM Filtrelerine Takılan Noktalar (Points Attached to SPAM Filters)

Format (Format): Tüm yazının büyük harf olması – parlak renkler kullanılması – gereğinden fazla ünlem kullanılması

İçerik (Content): "Para Almak"tan söz eden, geri ödeme, az ödeyin, daha az ödeyin, paranız iade vb. içerikler

Kod (Code): Özensiz yazılmış kodlar, fazla kullanılmış taglar, özellikle Word'den üretilmiş HTML kodları

Görseller ve İmajlar (Visuals and Images): Fazla ve sadece görsel kullanımı – İçerikte Yazı & Görsel dengesi olmaması. Kullanılan veri setinde bu tür noktaları tespit eden alanlar mevcuttur.

3.7. SPAM Filtrelerinin Kontrol Ettiği Alanlar (Spam Filter Controlled Areas)

Konu (Topic): SPAM filtreleri için anahtar konu ilişki ve ilgidir.

To Alanı (To Area): SPAM Filtreleri, gönderim yapılan alıcı ile tanınıp tanınmadığını kontrol eder. Örneğin, "…@….com" şeklindeki bir adrese mail atılmasındansa <İsim Soyisim> şeklinde, isimlere gönderim yapılması SPAM filtreleri için olumlu bir puanlamadır.

İçerik (Content): İçerikte tanıtma yazılarına, şirket hakkında bilgi içeriklerine yer gönderilen maillerde yakınlık fark edileceğinden olumlu bir puanlama sağlar. Ayrıca içerik temiz, dengeli, SPAM kelimelerden ve noktalama işaretlerinden uzak olmalıdır.

IP Adresi (IP Adress): SPAM filtreleri, gönderici IP'sinin kara listedeki IP alanlarından olup olmadığına dair kontrol yapar.

Kimden (From): Alıcılarla tanıdık olmak önemlidir. SPAM Filtreleri, gönderen ve alıcının adres defterlerinde tanımlı olup olmadığını kontrol eder. Doğrulanmış domainler yerine Google, Yahoo gibi ücretsiz anonim domainlerin kullanılması SPAM oranını arttırır.

Domain–Alan Adı (Domain Name): SPAM gönderenler genellikle fake–sahte hesap kullanıyorlar ve Kimlik Denetleme– Authentication yapmadan gönderim yapar [16,17].

4. GELIŞTIRILEN UZMAN SİSTEMİN YAPISI VE UYGULANMASI (STRUCTURE AND IMPLEMENTATION OF DEVELOPED EXPERT SYSTEM)

Bu çalışmada, SPAM filtrelemek için bir uzman sistem tasarlanmıştır. Çalışma SPAM filtreleme yazılımlarından içerik filtreleme yöntemine bir örnek olacak niteliktedir. Bu yöntemle, e-posta içerisinde kelimeler aranarak bu kelimelerin SPAM olup olmadığına karar verilmektedir.

4.1. Veri Seti (Data Set)

Veri kümesi, UCI makine öğrenmesi [12] veri tabanından alınmıştır. Veri kümesi Hewlett-Packard laboratuarından elde edilen 4601 elektronik postadan oluşmaktadır. E-postalardan 57 özellik elde edilmiştir. İlk 48 özellik elektronik mesajlardan elde edilen kelimelerin frekanslarını göstermektedir. Bununla birlikte 49-54 arasındaki 6 özellik ise elektronik mesajlarda geçen ';', '(', '[', '!', '\$' ve '\#' gibi karakterlerin frekanslarını göstermektedir. 55-57 arasındaki özellikler ise, büyük harflerle yazılmış kelimelerin "toplam harf sayısı", "ortalama harf sayısı" ve "en uzun kelimenin harf sayısı"nı belirtmektedir. 58. özellik ise elektronik postanın SPAM olup olmadığını belirtmektedir.

Veri setindeki 4601 e-postanın 1813 (%39.4) adedi SPAM, 2788 (%60.6%) adedi ise SPAM değildir. E-posta girişlerinde kullanılan kelime örnekleri: 'free', 'credit', 'business' (genellikle SPAM e-postalarda kullanılan) ve 'project', 'meeting' , 'George', 'hp' (genellikle SPAM olmayan e-postalarda bulunan)

Elektronik mesajlardan alınan kelime ve karakterler Çizelge 1'de verilmiştir.

Bazı giriş özellikleri SPAM e-postalar için yüksek değerler içermektedir. Örneğin, büyük harf içeren kelime

Çizelge 1. Veri setinde kullanılan özellikler (Features used in dataset)

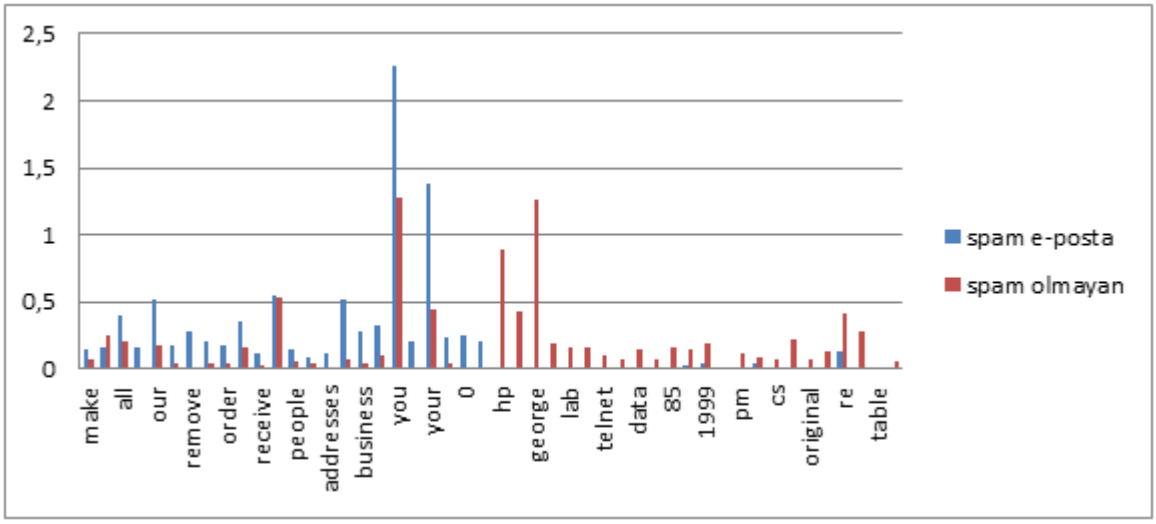
1-make	10-mail	19-you	28-650	37-1999	46-edu
2-address	11-receive	20-credit	29-lab	38-parts	47-table
3-all	12-will	21-your	30-labs	39-pm	48-conference
4-3d	13-people	22-font	31-telnet	40-direct	49-;
5-our	14-report	23-0	32-857	41-cs	50-(
6-over	15-adresses	24-money	33-data	42-meeting	51-[
7-remove	16-free	25-hp	34-415	43-original	52-!
8-internet	17-business	26-hpl	35-85	44-project	53-$
9-order	18-email	27-george	36-tecnology	45-re	54-#

4.2. Veri Setinin Azaltılması (Data Set Reduction)

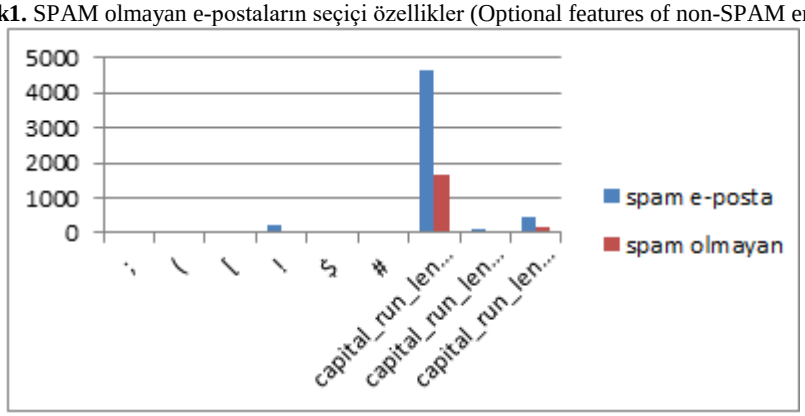
Veri seti incelendiğinde bazı giriş özelliklerin gereksiz ve yüksek varyansa sahip olmadığı görülmüştür. Bu durum, ayırt etme işleminde (tespit etme sürecinde) zorluk çıkaracaktır. Bu yüzden bazı özelliklerin değerlendirilmesine gerek kalmamıştır. Dolayısıyla, veri setinde gelen özelliklerden make, adress, all, 3d, our, internet, mail, people, report, adresses, business, e-mail, sayısı ve "!" işareti gibi alanlar. Bu alanlar SPAM e-posta belirlemek için kullanılabilir.

Aynı şekilde conference, George, you, 650 alanları kişisel veri içerdiğinden bunlar da SPAM olmayan e-postaların belirlenmesinde kullanılabilir.

Grafik 1'de SPAM olmayan e-postaların seçici nitelikleri gösterilmiştir. Grafik 2'de ise SPAM olan e-postaların seçici nitelikleri gösterilmiştir. Her iki grafikte uzman



Grafik1. SPAM olmayan e-postaların seçiçi özellikler (Optional features of non-SPAM emails)



Grafik 2. SPAM e-postaların özellikleri (SPAM email features)

font, (, [, # gibi özellikler uzman sistem için belirleyici özellik taşımadığından dikkate alınmamıştır. Ancak gerektiğinde yeni kurallar için seçilebileceklerdir.

sistem tasarımında daha önce sayılan öz niteliklerin neden elendiği açıkça görülmektedir.

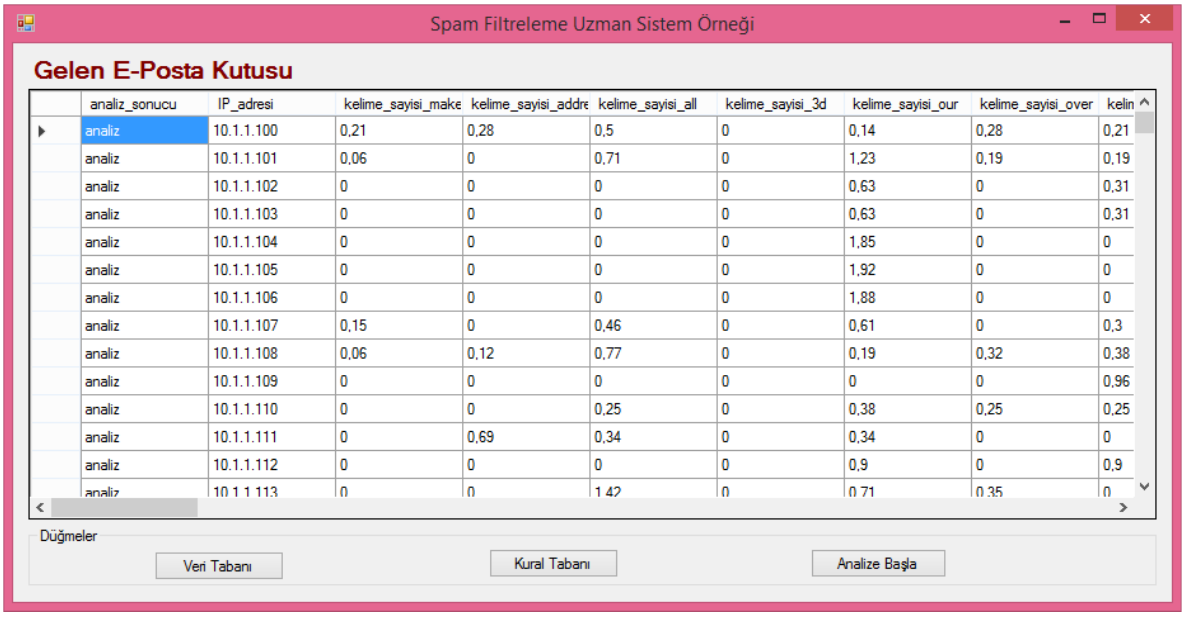
4.3. **Kullanıcı Arabirimi** (User Interface)

Kullanıcı arabirimi kullanıcı ile program arasında iletişimi sağlar. Kullanıcı, bu arabirim sayesinde bilgi tabanını kontrol etme, veritabanındaki özellikleri seçme, kuralları ekleme, çıkarma gibi işlemleri gerçekleştirebilir.

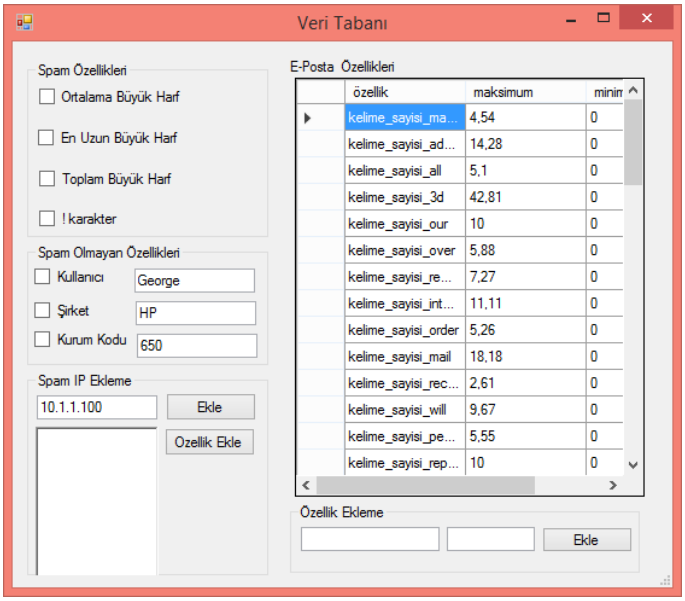
Geliştirilen uzman sistemin kullanıcı arabiriminin Ana Ekranı Şekil 2'de gösterilmektedir. Kullanıcı bu arayüzü kullanarak SPAM filtrelemesinde kullanılacak IP'ler ve özellikleri seçeceği "Veri Tabanı" bölümü ile kurallar ekleyebileceği "Kural Tabanı"na erişebilir. SPAM filtrelemesinde kullanacağı özellikler ile kurallar seçildikten sonra "Analize Başla" düğmesine basılarak sistem tarafından e-postaların SPAM olup olmadığına karar verilir.

4.4. **Bilgi Tabanı** (Konwledge Base)

Bilgi tabanı, problem hakkında ön bilgilerin bulunduğu 'veritabanı' ve uzman bilgisinin aktarıldığı 'kural tabanı'ndan oluşmaktadır. Yapılan çalışmada kullanıcı programın Şekil 3'te yer alan "Veri Tabanı" bölümünde e-postalarda özellik seçimi ile engellenmesini istediği e-posta IP adreslerini ekleme işlemini yapabilmektedir. Ayrıca kullanıcı isterse, "E-posta Özellikler" tablosundan özellik silebilir veya yeni özellik ekleyebilir. Böylece, bilgi tabanı güncellenerek kullanıcıya kendi seçimine göre gelen e-postaların SPAM olup olmadığına karar verilmesi sağlanmaktadır.



Şekil 2. **Geliştir**ilen uzman sistemin ana ekranı (Main screen of developed expert system)



Şekil 3. Veri tabanı ekranı (Database screen)

Bu bölümdeki özellikler SPAM e-postalarda bulunan tipik kelimelerin e-posta metninin içerisinde geçme oranına göre belirlenmiştir.

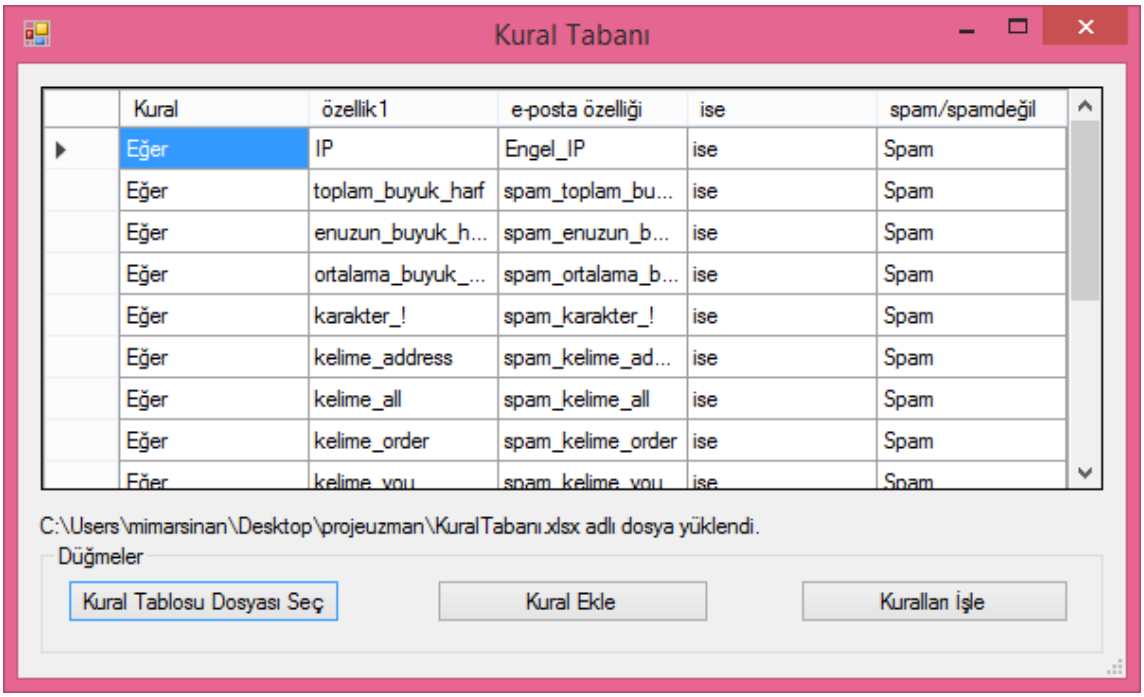
Kullanıcı, Şekil 4'te gösterilen Kural Tabanında e-postaların SPAM olup olmadığına karar veren kuralları seçme, silme, ekleme işlemlerini yapabilmektedir.

Kural Tabanında özelliklere göre oluşturulmuş 57 kural yer almaktadır. Kullanıcı "Kural İşle" düğmesine bastığında Veri Tabanı bölümünde seçilen özellikleri içeren kurallar aktif olacaktır. Bu durumu göstermek için program seçilen kuralları kırmızıya boyanmakta ve bir mesaj kutusuyla kullanıcıya bilgi verilmektedir. Sistemde kural tabanı, "eğer-ise" yapısında oluşturulmuştur.

Çıkarım mekanizması bilgi tabanındaki verileri ve kuralları kullanarak elde ettiği bu çıkarımları kullanıcıya aktarır [17]. Geliştirilen yazılımında çıkarım mekanizması kuralları yorumlamada, Çizelge 2'de gösterilen kurallar kullanıcının veri tabanından seçtiği özelliklerden başlayarak uygun kuralın bulunması durumunda ilgili kuralın mevcut şartlarını belirlemeyi temel alan ileri zincirleme metodu kullanılmıştır.

4.6. Açıklama Ünitesi (Description Unit)

Açıklama ünitesi, elde edilen sonuçların kullanıcılara iletilmek üzere uzman sistem tarafından raporlandığı bölümdür. Raporun doğruluğu geri bildirim için önemlidir [18]. Uygulamada açıklama bölümü, her bir e-postanın belirlenen kurallarla değerlendirilerek Şekil 5'te gösterildiği gibi açıklama bölümünde yazı ile e-postanın SPAM olduğunu kırmızı renkte, SPAM olmadığını sarı renkte kullanıcıya geri dönüş yapmaktadır



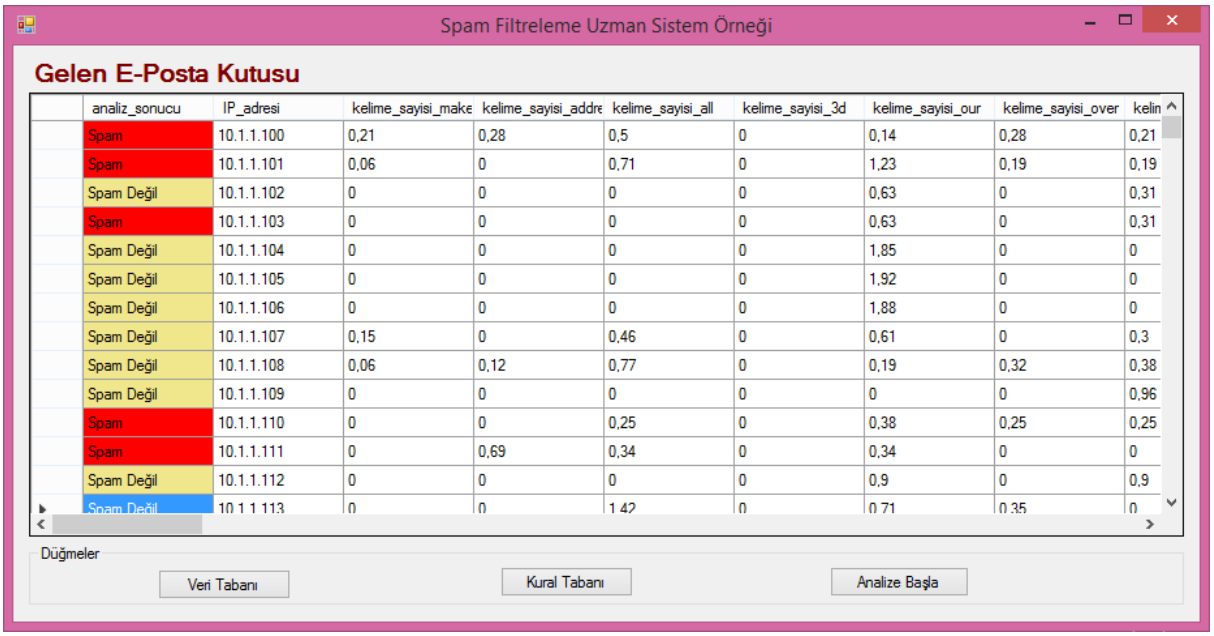
Şekil 4. Kural tabanı ekranı (Rule base screen)

4.5. Çıkarım Mekanizması (Inference Mechanism)

Çıkarım mekanizması, kurallar tabanındaki kuralları kullanarak anlamlı çıkarımların yapıldığı birimdir.

Çizelge 2. Çıkarım mekanizmasının kullandığı kuralların bir kısmı (Some of the rules used by the extraction mechanism)

Kural	Özellik	E-posta Özelliği	Sonuç	SPAM/SPAM Değil
Eğer	IP	Engel_IP	İse	SPAM
	toplam_buyuk_harf	SPAM_toplam_buyuk_harf		SPAM
	enuzun_buyuk_harf	SPAM_enuzun_buyuk_harf		SPAM
	ortalama_buyuk_harf	SPAM_ortalama_buyuk_harf		SPAM
	karakter_!	SPAM_karakter_!		SPAM
	kelime_address	SPAM_kelime_address		SPAM



analiz_sonucu	IP_adresi	kelime_sayisi_make	kelime_sayisi_addre	kelime_sayisi_all	kelime_sayisi_3d	kelime_sayisi_our	kelime_sayisi_over	kelin
Spam	10.1.1.100	0,21	0,28	0,5	0	0,14	0,28	0,21
Spam	10.1.1.101	0,06	0	0,71	0	1,23	0,19	0,19
Spam Değil	10.1.1.102	0	0	0	0	0,63	0	0,31
Spam	10.1.1.103	0	0	0	0	0,63	0	0,31
Spam Değil	10.1.1.104	0	0	0	0	1,85	0	0
Spam Değil	10.1.1.105	0	0	0	0	1,92	0	0
Spam Değil	10.1.1.106	0	0	0	0	1,88	0	0
Spam Değil	10.1.1.107	0,15	0	0,46	0	0,61	0	0,3
Spam Değil	10.1.1.108	0,06	0,12	0,77	0	0,19	0,32	0,38
Spam Değil	10.1.1.109	0	0	0	0	0	0	0,96
Spam	10.1.1.110	0	0	0,25	0	0,38	0,25	0,25
Spam	10.1.1.111	0	0,69	0,34	0	0,34	0	0
Spam Değil	10.1.1.112	0	0	0	0	0,9	0	0,9
Spam Değil	10.1.1.113	0	0	1,42	0	0,71	0,35	0

Şekil 5. Açıklama ekranı (Description screen)

5. SONUÇ VE ÖNERILER (CONCLUSIONS AND RECOMMENDATIONS)

Gündelik yaşantımızın bir parçası haline gelen elektronik posta trafiğinin büyük bir kısmını oluşturan SPAM elektronik postaları hem kullanıcılar için hem de internet trafiği için önemli bir sorun haline gelmiştir. Bu sorundan kurtulmak için kullanıcılar filtreleme yazılımları kullanmaktadır.

Bu çalışmada SPAM filtrelemesi için hem içeriğe göre hem de IP adreslerine göre filtreleme yapan bir uzman sistem tasarlanmıştır. SPAM'lar çok çeşitli olduğundan giriş özellikleri buna göre belirlenmelidir. Kullanıcı kendine uygun filtreleme seçenekleri ile gelen e-postaların SPAM olup olmadığını belirleyebilmektedir.

Çalışmada, daha önce farklı metotlar kullanılarak gerçekleştirilen SPAM filtreleme faaliyeti uzman sistemler tekniği kullanılarak gerçekleştirilmiştir. Böylece hem bu faaliyetler daha hızlı ve güvenilir bir şekilde gerçekleşmekte hem de harcanacak iş gücü azalmaktadır.

KAYNAKLAR (REFERENCES)

[1] Kaya, Y., Yeşilova, A., & Tekin, R., "İstenmeyen Elektronik Postaların (Spam) Filtrelenmesinde Kaba Küme Yaklaşımının Kullanılması", **Elektrik-Elektronik Bilgisayar Sempozyumu** (FEEB 2011), Elazığ, 148-153, (2011).

[2] Sivanadyan, T., "Detecting Spam emails using neural networks", www.cae.wisc.edu/~ece539/project/f03/sivanandyan.pdf (Erişim Tarihi: 12.04.2016)

[3] Gündüz, H. C., "Spam 2.0, Tespit ve Engelleme Yöntemleri". ***Akademik Bilişim 07***, Kütahya, 677-683, (2007).

[4] Nabiyev, V. V., "Yapay zeka: problemler-yöntemler-algoritmalar". ***Seçkin Yayıncılık***, Ankara, (2005).

[5] A. Doğaç, "Uzman Sistemler", ***Elektronik Mühendisliği Dergisi***, 373: 87-91, (1990).

[6] Üstkan, S., "Uzman Sistemler-Genel", ***Sakarya Üniversitesi***, (2007).

[7] Calp, M. H., & Şahin, İ. The determination by using fuzzy expert system of the usability level of website user interface design. ***International Journal of Human Science***, Volume: 10 Special Issue, 35-44.

[8] Baykal, N., & Beyan, T., "Bulanık mantık: uzman sistemler ve denetleyiciler", ***Bıçaklar Kitabevi***, Ankara, (2004).

[9] Şahin, İ., Calp, M. H., & Özkan, A. An Expert System Design and Application for Hydroponics Greenhouse Systems. ***Gazi University Journal of Science***, *27*(2), 809-822, (2014).

[10] Şahin, İ., Calp, M. H., & Sönmez, A., Elektronik Cihazlarda Arıza Teşhisi İçin Bir Uzman Sistem Uygulaması. ***Selçuk-Teknik Dergisi***, *11*(1), 8-18, (2012).

[11] Bilişim Teknolojileri., "İnternet ve E-Posta Yönetimi", ***MEB***, Ankara, (2011).

[12] İnternet: http://yunus.hacettepe.edu.tr/~sadi/dersler/e-posta_kurallar.html, (Erişim Tarihi: 15.05.2016)

[13] İnternet: https://tr.wikipedia.org/wki/E-posta , (Erişim Tarihi: 15.05.2016)

[14] İnternet:https://seminer.linux.org.tr/seminer%adnotlari/spamiltreleme.sxi, (Erişim Tarihi: 15.05.2016)

[15] Şahinaslan, Ö, Borandağ, E, Can, E, Şahinaslan, E., "Posta Sunucularında Spam Önleme Teknikleri", ***Akademik Bilişim'09***, Şanlıurfa, 737-743, (2009).

[16] Yüksel, M. E, Odabaşı, Ş. D., "SMTP Protokolü ve Spam Mail Problemi", ***Akademik Bilişim 2010***, Muğla, (2010).

[17] Erkalan, M, Calp, M. H, Şahin, İ., "Çoklu Zekâ Kuramından Yararlanılarak Meslek Seçiminde Kullanılacak Bir Uzman Sistem Tasarımı ve Gerçekleştirilmesi", ***International Journal of Informatics Technologies***, 5 (2): 49-55, (2012).

[18] Şahin, İ, Calp, M. H, Akça, Ö., "Kredibilite Notu Değerlendirmeye Yönelik Bir Uzman Sistem Yaklaşımı". ***Politeknik Dergisi***,14 (1): 79-83, (2011).