



Yayın Geliş Tarihi: 01.02.2018  
Yayına Kabul Tarihi: 08.05.2018  
Online Yayın Tarihi: 26.10.2018

Cilt:4, Sayı:1, Yıl:2018, Sayfa 1 -14  
ISSN: 2148-3752

## İNSAN OMURGASI VERİ SETİNİN SINIR-BULANIK SINIFLAYICI İLE ÖZNİTELİK TESPİTİ VE SINIFLANDIRILMASI

Sümeyye ÇELİK  
Özlem ÇETİNKAYA BOZKURT  
Melike ŞİŞECİ ÇEŞMELİ

**Öz:** Tıp alanında, insan omurgasının sağlıklı olup olmadığına karar verilmesi için kontrol edilen bazı değerler belirlenmiştir. Hekim, teşhis koyarken bu değerlerden faydalanır. Veri madenciliği yöntemleri ile hekimlere yardımcı olacak sistemler geliştirilebilir. Kullanılan sistemlerde amaç doğru yöntemleri seçmek ve teşhisin doğruluk oranını arttırmaktır. Bu araştırmada kullanılan veri seti 2011 yılında Brezilya'nın Ceará eyaletindeki Fortaleza şehrinde bulunan Monte Klinikum hastanesinde yapılan bilimsel bir araştırmanın sonuçlarıdır ve UCI machine learning repository sitesinden temin edilmiştir. Amaç bireyleri hasta olup olmadıklarına göre iki kategoride sınıflandırmaktır. Benzer çalışmalarda sınıflandırma işlemi yapıldıysa da bu çalışmada önemli öznelik seçimi üzerinde çalışılmıştır. Bu özneliklerin önem derecesinin artırılmasıyla daha yüksek bir başarı sağlanması amaçlanmıştır. Veri seti üzerinde başarıyı artırmak için birçok yöntem denenmiş ve istenilen başarı artışı elde edilmiştir. Uyarlamalı Sinir-Bulanık Sınıflayıcı yöntemi ile test verisinde %83.7419 'den %96.1935 oranına ulaşan bir başarı artışı sağlanmıştır.

**Anahtar kelimeler:** Omurga, Sinir-Bulanık Sınıflayıcı, Öznelik Seçimi, Sınıflandırma

## SELECTION OF ATTRIBUTES AND CLASSIFICATION OF VERTEBRAL COLUMN DATA SET WITH NEURO-FUZZY CLASSIFIER

**Abstract:** In medicine, certain parameters, checked in order to decide if vertebral column is healthy or not, are established. Physicians make use of these parameters while diagnosing. Systems which help doctors can be developed with data mining methods. The purpose of the used systems is to select correct methods and increase the accuracy rate of the diagnostic. The data set used in this research are the results of a scientific research conducted in Monte Klinikum Hospital, located in Fortaleza city, in Brazil's Ceara state. And they are procured from the UCI machine learning repository site. The aim is to classify persons in two categories according to whether they are ill or not. Although classification was done in similar studies, important attribute selection was studied in this study. It is aimed to achieve a higher success by increasing importance of these attributes, and achievement of desired success has been achieved With the Nerve-Fuzzy Classifier, a success rate ranging from 83.7419% to 96.1935% was achieved in the test data.

**Keywords :** Vertebral Column, Neuro-Fuzzy Classifier, Feature Selection, Classification

## **GİRİŞ**

Bilişim teknolojilerinin gelişmesi, hazır veri toplama algoritmaları ve depolama maliyetlerinin azalmasıyla birlikte her geçen gün artan veri miktarı, büyük veri yığınlarına sebep olmaktadır. Oluşan yüksek miktardaki bu veriler arasındaki ilişkileri ortaya çıkararak verileri anlamlı bilgilere dönüştürmek gelişen ülkeler, kurum ve kuruluşlar için oldukça önemlidir. Bu da teknolojinin etkin ve verimli kullanılmasına bağlıdır. Bu gelişmelerden her alanda olduğu gibi sağlık alanında da faydalanılmaktadır. Sağlık alanındaki veriler, hastanın tüm yaşamını etkileyebilecek kararlar alınmasına sebep olabilmektedir. Bu sebeple, yapılacak çalışmalar önem kazanmaktadır.

Tıp bilimi insanların sağlık sorunlarıyla ilgilenmektedir. Çalışma alanının insan hayatı olması sebebi ile doğru teşhisin konulması hızlı olmasının yanı sıra oldukça zor, karmaşıktır. Teşhis konulurken verilecek yanlış bir karar telafisi mümkün olmayan sonuçlara sebep olabilir çünkü İnsan hayatı hiçbir şeyle geri getirilemeyecek kadar önemlidir. Bu sebeple hekime doğru teşhisi koymasında yine hekimin de içinde bulunduğu ve teknoloji ile harmanlanarak oluşturulan bilgisayar sistemlerine ihtiyaç vardır. Bu sistemler hekimi ortadan kaldırmaya yönelik değil, hekime teşhisin hızlandırılması konusunda yardımcı olacak sistemlerdir. Bilişim ile oluşturulan birçok sistem ile hekimlere teşhis için zaman kazandırılırken, tedavi ve daha spesifik alanlara yoğunlaşarak daha doğru kararlar verebilmelerine olanak sağlanmaktadır. Aynı zamanda hekimin üstündeki gereksiz iş yoğunluğunun da alınmasıyla birlikte verim arttırılabilmektedir (Altıntaş, 2010).

Hekime doğru kararı vermesinde yardımcı olacak sistemler oluşturulurken uzman bilgisinden, önceki hastalık teşhislerindeki verilerden ve bilişim teknolojilerinden yararlanmak gerekir. Girilen verilerin aynı platformda toplanmıyor olması bir takım analizleri zorlaştırır (Akal, 2016). Mevcut veriler, veri ön işleme gibi veri madenciliği yöntemleri kullanılarak ortak bir platformda bir araya getirilebilir. Önceki verilerden anlamlı bilgi ve örüntüler çıkarılabilir. Bu örüntüler yardımcı ile sonuçlar daha doğru tahmin edilebilir. Bu noktada veri madenciliği yöntemleri çok önemli bir paya sahiptir. Veri madenciliği, yapay zeka, makine öğrenmesi gibi teknikler sayesinde sonuca etki değeri daha yüksek öznelilikler belirlenebilir. Böylece doğru noktalara yoğunlaşılabilir. Gözden kaçan bir değer daha kolay tespit edilebilir (Reddy vd., 2012).

İnsan beyni çeşitli yollarla öğrenme gerçekleştirir ve öğrendiği bu bilgilerle nihai kararlar verebilir. Fakat bu kararlarına o andaki psikolojik ve zihinsel durumu, yorgunluk, dikkat dağınıklığı, bilginin uzun zamandır kullanılmıyor olması, öncelik durumu gibi sebepler etki edebilmekte ve kişinin yanlış ya da eksik karar vermesine sebep olabilmektedir (Çetinkaya Bozkurt vd., 2016). Bu açıdan bakıldığında hekim başkasının sağlık durumu ile ilgili bir karar vermek durumundadır ve yanlış bir karar vermesi çok büyük hatalara yol açabilir. Eğer bu noktada uzman bilgisi ve teknoloji sayesinde bu bilgiler makinelere öğretilirse hekimin bireysel hatalarının önüne geçilmesine de yardımcı olunmuş olur (Budak ve Bozkurt., 2013).

Model çıkarımı ve veri madenciliği yöntemlerinde öznitelik seçimi çok önemlidir. Özellikle büyük verilerle çalışılıyor ise bu durum daha da önemli bir hale gelmektedir çünkü çok veri ile çalışmak uzun, maliyetli ve zahmetli bir süreçtir. Bu süreçte kullanılan özniteliklerin sonuca etki değeri tespit edilebilir ise, sonuca etki etmeyen öznitelikler çıkarılarak yapılan analizde veri sayısı azalacağından daha hızlı sonuçlar alınabilir (Çetişli, 2006). Bu aynı zamanda bellekten de tasarruf sağlanması demektir (Karasulu, 2016). Bunun yanında sonuca etki değeri fazla olan özniteliklerin tespit edilmesi ile birlikte daha anlamlı yorumlar yapılabilir ve önemli görülen noktalara odaklanılarak daha başarılı sonuçlar elde edilebilir (Çetişli, 2006).

Sağlık alanında veri madenciliği yöntemleri literatürde çokça rastlanmaktadır. Bu yöntemler geliştirilirken alanında uzmanından alınan teknik bilgilere ihtiyaç duyulur. Örneğin omurga ve pelvisin sagittal yerleşimini tanımlamak için çeşitli parametreler oluşturulmuştur. Yetişkinlerde ve çocuklarda sagittal sırt ve spinopelvik parametreler farklı olmakla birlikte, bu parametrelerin küresel dengeyi korumak için birbirleriyle ilişkili olduğu belirlenmiştir (da Rocha Neto vd., 2009; da Rocha Neto vd., 2011). Bu ilişkiler ortaya konularak omurganın durumunun normal olup olmadığı durumu tespit edilebilir (Ghandhari vd., 2013). Aynı zamanda sonucu etkileyen özniteliklerin önem derecesi çıkarılabilir.

Bu çalışmada omurga durumunun normal olup olmadığına karar verilmesi için etki değeri daha fazla olan özniteliklerin seçilmesi konusuna odaklanılmıştır. Bunun için tıp literatüründe kabul görmüş olan Pelvic incidence, Pelvic tilt, Lumbar lordosis angle, Sacral slope, Pelvic Radius ve Degree spondylolisthesis şeklindeki öznitelikler kullanılmıştır. Bu değerler UCI machine learning repository sitesinden alınan Vertebral Column veri kümesinden temin edilmiştir (Machine Learning Repository, <https://archive.ics.uci.edu/ml/datasets/Vertebral+Column>, 21.12.2016) . USPS yöntemi kullanılarak bu öznitelikler arasından sınıflandırmada önem derecesi fazla olan öznitelikler tespit edilmiş ve önemli olmadığı tespit edilen öznitelikler devre dışı bırakılarak bir sınıflandırma işlemi gerçekleştirilmiştir. Böylece daha az öznitelik ile daha doğru bir sınıflandırma yapılmıştır.

## **VERİ MADENCİLİĞİ**

Mevcut olarak elde bulunan fakat işlenmediği zaman bir anlam ifade etmeyen, karmaşık verilerin bir sistematik içerisinde işlenmesi sürecidir. Bu süreç sonucunda yığınlar içinde gizlenmiş faydalı olan bilgiler elde edilebilir (Albayrak ve Koltan Yılmaz, 2009). Böylece geçmişteki tecrübelerden faydalanılarak gelecekle ilgili daha doğru kararların alınmasına yardımcı olunabilir. Veri madenciliği yöntemleri ile veri kümesi içerisinde ki keşfedilmemiş örüntüler ortaya çıkarılır (Hong ve Wu, 2011). Çıkarılan bu örüntüler sayesinde veriler anlamlandırılarak bilgiye dönüştürülür. Bu süreçte, problemin türüne göre hali hazırda geliştirilmiş olan analiz yöntemleri kullanılabilir. Var olan yöntemlerin yetersiz kaldığı noktalarda ise yeni yöntemler geliştirilebilir (Koyuncugil ve Özgülbaş, 2009). Problem tanımlama, veri anlama, veri önileme, veri temizleme, modelleme, değerlendirme, yaygınlaştırma şeklinde

aşamalardan meydana gelmektedir (Kızılkaya Aydoğan vd., 2007). Veri madenciliği: sınıflandırma, kümeleme, zaman serileri ve sıralı model araştırması şeklinde önemli ilişkilendirme kuralları içerir (Gharib vd., 2010).

### **Veri Madenciliğinde Öznitelik Seçimi**

Öznitelik seçimi veri madenciliği yöntemleri içinde önemli bir paya sahiptir (Çetişli, 2016). Girdi öznitelikleri içerisinden çıkarılması sonucu değiştirmeyecek olanlar var ise onların tespiti işlemidir. Önem derecesi yüksek olan öznitelikler kullanılarak daha iyi sonuç elde etme performansına sahip bir öznitelik alt kümesi bulunmuş olacaktır (Altındağ, 2013).

Örneğin, Pençe ve Tarhan (2016) öğrenim sürelerini uzatan öğrencilerin tahmin edilmesi üzerinde bir çalışma yapmışlardır. Bu çalışmada aynı zamanda sonuca etki değeri yüksek olan öz niteliklerin seçilmesi işlemi de gerçekleştirilmiştir. Öğrencilere ait cinsiyet, yaş, fakülte, sınıf gibi toplam da 9 öznitelikten yararlanılmıştır. Bu çıkarımların gerçekleştirilmek için yapay sinir ağları (YSA) ve USBS gibi yöntemler kullanılmıştır.

Gündüz ve arkadaşları (2013), Kalabalık dinamiklerin çıkarılması için özniteliklerin tespiti ve takibi için bir çalışma yapmışlardır. Özniteliklerin tespitinin yanı sıra uygulanabilirlik de incelenmiştir. Bunun için literatürde kabul edilen bazı yöntemler kullanmış ve öznitelik seçimine vurgu yapılmıştır. Sonuç olarak daha hızlı olan algoritma tespit edilmiştir.

Onan ve Korukoğlu (2016), Bu çalışmada dokuz farklı metin verisi sınıflandırılmıştır. Bu sınıflandırmalar için korelasyon tabanlı (CS) ve tutarlılık tabanlı (CBS) gibi öznitelik seçim yöntemleri ile elde edilen sonuçlar değerlendirilmiştir. Sonuç olarak en yüksek başarı oranı naive bayes ve sıra arama algoritmasının birlikte kullanılması ile elde edilmiştir.

Gökberk ve arkadaşları (2003), Çalışmalarında altküme seçim yöntemleri ile öznitelik tespiti yapmayı hedeflemişlerdir. Sonuç olarak en yüksek başarı gezen ileri seçim (SFFS) ve genetik algoritmanın birlikte kullanılması ile sağlanmıştır.

### **Tıp Alanında Veri Madenciliği Kullanımı**

Hastalık teşhisi konulurken vakanın ciddiyet durumu arttıkça verilen kararın önem derecesi artar ve sürecin hızlandırılması gerekir. Kısa zamanda doğru karar verilebilmesi için Tıp alanında Veri Madenciliği yöntemlerinden faydalanmak gerekir (Reddy vd., 2012). Bu konuda literatür de birçok araştırma yapılmıştır. Örneğin şekil ve yön öznitelikleri kullanarak omurga ve pelvisin sagittal dengesinin analizi yapılmıştır. Bu çalışma pelvis ve servikal, torasik ve lomber omurganın şekil ve yöneliminden elde edilen indeksleri kullanarak insan gövdesinin küresel sagittal dengesini analiz etmek ve karakterize etmek için bir yöntem sunmaktadırlar. Yöntem: 160 asemptomatik genç erişkin gönüllüden oluşan bireylerin yanıl röntgen verileri kullanılarak gerçekleştirilmiştir. Modeller, bir bilgisayar yazılımı kullanarak yaratılmıştır ve anatomik seviyeler hesaplanmıştır (Berthonnaud vd., 2005).

da Rocha Neto ve arkadaşları (2009), Ortopedi alanında tıbbi bir tanı sorununa uygulandığında, bağımsız makine öğrenme algoritmaları (SVM, MLP ve GRNN) ve sınıflandırmalardaki kombinasyonları arasındaki performans karşılaştırmasının sonuçlarını rapor etmektedir.

Sonuç olarak olağandışı değerler veride bulunmadığında (yani P =% 0) sınıflandırıcının performansının maksimum olduğu açıktır.

da Rocha Neto ve arkadaşları (2011) Bu çalışmada, vertebral kolonda patolojilerin teşhisi konusundaki kararın alınması için çeşitli teknolojik tekniklerle yapılan bir karşılaştırma gerçekleştirilmiştir. Sonuç olarak ise SVM ve RejoSVM yöntemleri karşılaştırılmıştır ve RejoSVM'nin daha doğru sonuçlar verdiği ortaya konulmuştur. .

Chae ve arkadaşları (2001), Kore sağlık sigortası kurumu verileri üzerinde bir çalışma yapmışlardır. Çalışmanın amacı belirlenen öznitelikler yardımı ile bireylerin yüksek tansiyonunun olup olmadığının tahminlenmesi olmuştur. Logistic Regression, chaid, c 5.0 gibi bir çok tahmin modeli kullanılmıştır. Sonucunda ise yüksek tansiyona etki değeri fazla ve az olan öznitelikler belirlenmiştir.

Demircioğlu ve Bilge (2015), Yapılan bu çalışmada gelişen gen araştırmaları ile birlikte çok fazla bilgi üretildiği ve bu bilgilerin doğru şekilde analiz edilmesi gerektiği konusuna değinilmiştir. Makalede veri madenciliği yöntemleri kullanılarak kanser üzerine yapılan çalışmalar incelenmiştir. Uygulama kısmında özniteliklerin belirlenmesi için iki yöntem kullanılmıştır. Önemli görülen öznitelikler belirlendikten sonra, genler üzerinde sınıflandırma yapılmıştır ve sonuçlar karşılaştırılmıştır.

İşler ve Narin (2012), kalp yetmezliği hastalığına sahip bireyleri teşhis etmeye çalışmışlardır. Aynı algoritma üzerinde küme sayıları belirlenmeye çalışılmıştır. En yüksek başarının, küme sayısı dört seçildiğinde sağlandığı görülmüştür.

Ganzert ve arkadaşları (2002) çalışmalarında, Yoğun bakım tıbbındaki solunum basıncı hacim eğrilerini indüktif makine öğrenimi ile analiz etmişlerdir. Çalışmada yetişkin solunum sıkıntısı sendromundan (ARDS) şüphe edilen havalandırılmalı hastalarda basınç-hacim eğrilerinin ölçülmesinde kullanılan farklı yöntemler karşılaştırılmıştır. Yöntem olarak endüktif makine öğrenme aracı olan C5.0 kullanılmıştır. Araştırmada sonuç olarak bu tahmin edilen eğrilere dayanarak, hastanın invazivliğini en aza indirgeyeceği düşünülen yöntemin seçilebileceği söylenmiştir. Aynı zamanda bu yönde daha çok araştırma yapılması gerektiği vurgusu yapılmıştır

Coulter ve arkadaşları (2007), Uluslararası farmakovijilansda antipsikotik ilaçlar ve kalp kası bozukluğu üzerine veri madenciliği çalışması yapılmışlardır. Yöntem olarak bayesian yönteminikullanmışlardır. Sonuç olarak bazı antipsikotik ilaçların kardiyomiyopati ve miyokardit ile ilişkili olduğu sonucunu önermişlerdir.

Göreke ve arkadaşları (2014), meme kanserini teşhis etmeye yardımcı olabilmek için bilgisayar sistemlerini kullanmayı hedeflemişlerdir. Bunun için teşhiste kullanılabilecek öznitelikleri belirlemeye çalışmışlardır. Belirlenen özniteliklerle yapılan test aşamasında yüksek derecede başarı elde edilmiştir.

## **ARAŞTIRMADA KULLANILAN YÖNTEM VE ÖZNETELİKLER**

Araştırmada kullanılan veriler 2011 yılında Brezilya'nın Ceará eyaletindeki Fortaleza şehrinde bulunan Monte Klinikum hastanesinde yapılan bilimsel bir araştırmanın sonuçlarıdır. Söz konusu araştırma sonucunda elde edilen bulgular daha sonraki çalışmalara destek olmak için

<https://archive.ics.uci.edu> internet sitesinde yayınlanmıştır. Bu çalışmada da Machine Learning Repository, <https://archive.ics.uci.edu/ml/datasets/Vertebral+Column> (21.12.2016) sitesinde yayınlanmış olan veri seti kullanılmış ve önceki çalışmalar baz alınarak eksik görülen bazı noktalara değinilmiştir. Veri setinin özellikleri şu şekildedir; Veriler birbiri ile alakalı olarak iki farklı sınıflandırma halinde düzenlenmiştir. Birinci sınıflandırmada hastalar, normal (100 hasta), Disk Hernia (60 hasta) veya Spondilolistez (150 hasta) olmak üzere 3 sınıfa ayrılmıştır. İkinci sınıflandırmada Disk Hernisi ve Spondilolistez kategorileri, 'anormal' olarak etiketlenen tek bir kategoride birleştirilmiştir. Bu nedenle ikinci amaç hastaları iki kategoriden birine ait olarak sınıflandırmaktır. Yani hasta olup olmadıklarına karar verilmesi amaçlanmaktadır. Bu sınıflandırmada ise kategoriler Normal (100 hasta) veya Anormal (210 hasta) şeklinde olmuştur. Yapılacak olan sınıflandırma da kullanılan öznitelikler Pelvic incidence, Pelvic tilt, Lumbar lordosis angle, Sacral slope, Pelvic Radius ve Degree spondylolisthesis şeklindedir. Toplamda 310 hasta bilgisi bulunmaktadır ve kayıp değer yoktur. Girilen hasta verilerinin öznitelik değerleri numerik, çıktı parametreleri ise kategorik şekildedir (Reddy vd., 2012). Bu kategorilerin belirlenmesi için Uyarlamalı Sinir-Bulanık Sınıflayıcı (USBS) yöntemi kullanılmıştır (Çetişli, 2006). USBS yöntemi sınıflandırma yapmanın yanı sıra yapılan bu sınıflandırmayı daha doğru yapmak için, sınıflandırma işleminden önce önemli öznitelikleri belirler ve önemli görülmeyen öznitelikleri sınıflandırma yaparken devre dışı bırakır. Dolayısı ile bu çalışmada USBS yönteminin kullanılması ile omurganın normal olup olmadığı konusunda teşhis konulurken omurganın durumunu etkileyen özniteliklerden hangisinin daha fazla etkili olduğu, hangisinin daha az etkili olduğu ve hatta hangi özneliğin etkisinin olmadığı tespit edilmiştir. Bu önem dereceleri USBS yöntemi içerisinde bulunan bulanık çıkarsama sayesinde tespit edildikten sonra yine USBS yöntemi ile sınıflandırma yapılmış ve ne kadar başarılı olduğu test edilmiştir. USBS yöntemi bulanık çıkarsama ile öznitelik belirleme yaptıktan sonra belirlenen bu öznitelikleri sınıflandırırken Yapay Sinir Ağı (YSA) tekniğini kullanır. Yani sınıflandırma yapılması için eldeki özniteliklere göre YSA eğitilir ve bu eğitim sonrasında, test verisi YSA tarafında sınıflandırılmış olur (Çetişli, 2006).

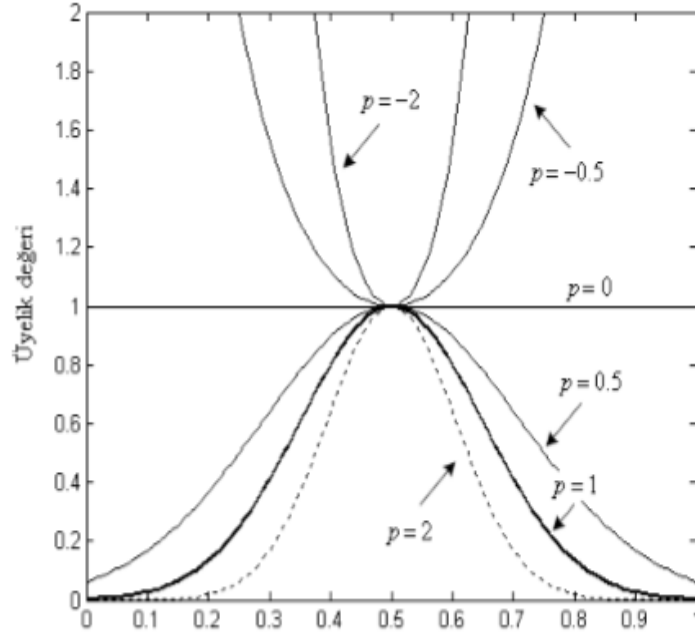
USBS temelde sınıflandırma yapan bir algoritmadır. Bu algoritmayı güçlü kılan ve diğer benzer algoritmalarından ayıran özelliği ise veri kümesi içindeki bütün özniteliklere bakarak, sonuç değişkeni üzerindeki önem derecelerini belirlemesidir (Pençe vd, 2013). Yani iki temel aşaması vardır; ilk olarak öznitelik tespiti yapar, ikinci olarak sınıflandırma işlemini gerçekleştirir. İkinci aşamada sınıflandırma yaparken tüm öznitelikleri kullanmaz. Birinci aşamada etkili olmadığı tespit edilen öznitelikler, verileri kümelere ayırırken kullanmamış olur. Yani kümeleri birbirinden ayırtırmayan, belirgin bir özelliğe sahip olmayan bir öznitelik devre dışı bırakılmış olur. Önem düzeyi yüksek olan öznitelikler belirlenir. Böylece daha az ve öz, belirleyiciliği yüksek olan öznitelikler kullanılmış olur. Daha az veri ile çalışıldığı için hız ve verimlilik artarken, bellekten tasarruf sağlanmış ve maliyetler azaltılmış olur. Test verisinde yüzde kaç oranında doğru sınıflandırdığı

görülebilmektedir. Eğer önemli öznitelikler belirlendikten sonra daha doğru bir sınıflandırma yapılmış olursa, sonucu etkileyen özniteliklerin doğru tespit edildiği düşünülebilir. Çünkü gerçek hayattaki problemlere de bakıldığında, bir problemi çözmek için her çözüm aynı oranda etkili olmayabilir. Bu yüzden önemli olan öznitelikleri tespit etmek önemlidir (Çetişli, 2006).

USBS, özniteliklerin önem derecesini belirlerken bulanık kuralları kullanır. Öznitelikler bu kurallar ile aşamalı olarak belirlenir. Yani bir özneliğin önemli olup olmadığına tek bir lojik kontrolle evet ya da hayır şeklinde cevap vermez. Bunun yerine özniteliklere "fazla", "çok fazla" ya da "az" şeklinde nitelermelerde bulunur. Böylece insan çıkarımlarına benzeyecek yapılar oluşturulmuş olur. Özniteliklerin önem düzeyi daha detaylı incelenerek belirlendiği ve insan çıkarsamalarına daha yakın olduğu için, daha doğru tespitlerde bulunulabilir. Öznitelikler belirlendikten sonra dilsel kuvvetleri kullanarak eğitilen YSA kullanılarak, sınıflandırma işlemi gerçekleştirilmiş olur. USBS sinir ağları kullanarak bir öğrenme gerçekleştirmesinin yanında bulanık mantığı da kullanan hibrit bir yapıya sahiptir. Özniteliklerin önem derecesi dilsel kuvvetler yardımı ile yapılır (Fırat, 2008).

A dilsel terimine uygulanan dilsel kuvvetler değişik  $p$  değerleri  $\{-2, -0.5, 0, 0.5, 1, 2\}$  için hesaplanan kuvvetlendirilmiş A dilsel terimleri Şekil 1. 'de görülmektedir. Şekil 1.'de ifade edildiği gibi,  $p < 0$  olması durumunda üyelik fonksiyonunun yönü değişmekte ve bunun sonucu olarak da hesaplanan üyelik değerleri 1'den büyük olmaktadır. Üyelik değerinin  $[0 + 1]$  aralığının dışına çıkması pek istenilmez. Çünkü bu aralığın dışına çıkılır ise, elde edilen kümeler bulanık olmaz. Bu sebepten dolayı  $p \geq 0$  olmalıdır. USBS yönteminde bu  $p$  değerlerine bakılarak özniteliklere önem düzeyleri atanır.  $P$  değeri sıfırdan bire doğru arttıkça özneliğin önem derecesi arttırılmış olur. Eğer  $p$  değeri sıfır çıkarsa öznitelik ile sonuç arasında bir ilişki olmadığı ortaya konulmuş olur (Çetişli, 2006).

Şekil 1 : A dilsel terimine uygulanan dilsel kuvvetlerin farklı  $p$  değerleri için yorumları



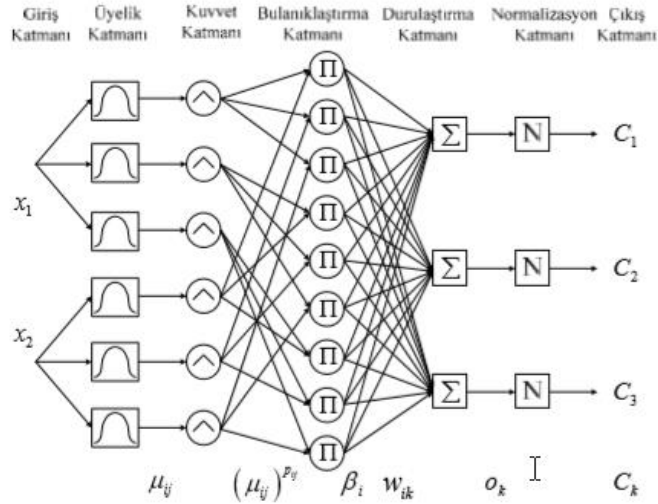
(Çetişli, 2006).

Sonuç olarak USBS girilen verileri sınıflandırmak için dilsel kuvvetleri kullanarak hangi özniteliğin daha fazla önem düzeyine sahip olduğuna karar verir. USBS'in çalışma mantığı "EĞER  $X_1 A_1^{p1}$  ve  $X_2 A_2^{p2}$  ise nesne C1 sınıfındadır" şeklinde formülize edilmiştir. Formüldeki X'ler öznitelikleri temsil etmektedir. A değerleri üyelik fonksiyonunu oluşturan dilsel kuvvetlerin matematiksel ifadeleridir. Sınıf etiketi olan C1 de nesnenin bulunduğu sınıfı ifade etmektedir.

USBS, doğru sınıflandırma için eğitilirken literatürde çokça kullanılan Yapay Sinir Ağlarını kullanır. YSA eğitilirken şekil 2 de gösterildiği gibi giriş katmanı, üyelik katmanı, kuvvet katmanı, bulanıklaştırma katmanı, doğrulama katmanı, normalizasyon katmanı ve çıkış katmanı şeklinde birçok katmandan oluşmaktadır.

Şekil 2 YSA kullanım şekli





(Çetişli, 2006).

Omurganın normal olup olmadığını teşhis ederken kullanılan Özniteliklerin özellikleri ile alakalı detaylı bilgi tablo 1 de gösterilmiştir.

Tablo 1 : Öznitelikler ve özellikleri

Öznitelik numarası	Öznitelik adı	Değer türü	Min Değerler	Max Değerler
1	Pelvic incidence	numeric	26,14792	129,83400
2	Pelvic tilt	numeric	-6,55495	49,43186
3	Lumbar lordosis angle	numeric	14,00000	125,74240
4	Sacral slope	numeric	13,36693	121,42960
5	Pelvic radius	numeric	70,08257	163,07100
6	Degree spondylolisthesis	numeric	11,05820	418,54310

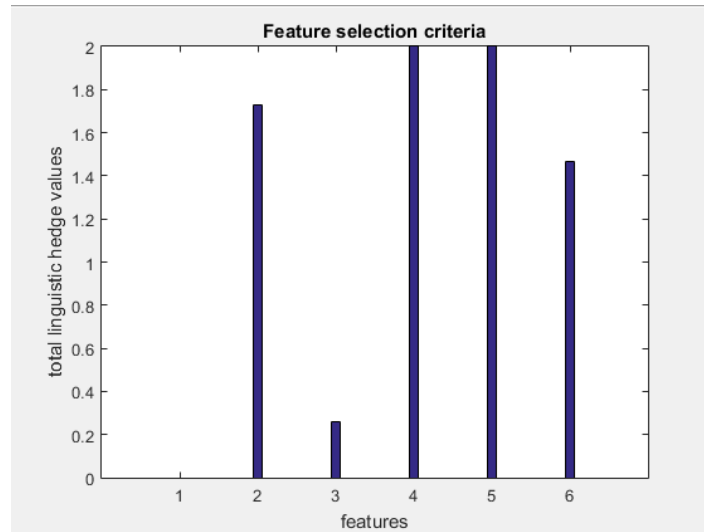
Kaynak: Machine Learning Repository, <https://archive.ics.uci.edu/ml/datasets/Vertebral+Column>, (21.12.2016).

Bir öznitelik bazı sınıflandırmalarda çok önemli olabilirken, başka bir sınıflandırmada hiçbir etkiye sahip olmayabilir (Pençe ve Çetişli, 2013). Örneğin yaş bazı araştırmalar için çok önemli olabilirken başka bir araştırma için daha az önemli olabilmektedir. Bu sebepten dolayı sınıflandırma işleminde her öz niteliğin farklı bir önem derecesi olduğu düşünülerek sınıflandırma yapılırsa daha başarılı sonuçlar elde edilebilir. Böylece kümeleri birbirinden ayırt edici özniteliklerin seçilmesi kolaylaşır. Gereksiz olduğu tespit edilen öznitelikler ise çalışmadan çıkarılabilir. Sınıflandırma için önem derecesi fazla olan özniteliklerin etkisi artırılarak doğru sonuçlar ortaya konulabilir. USBS de bu öz niteliklerin önem derecesini belirlemek için sinir ağlarının öğrenme şekli ile bulanık mantığın çıkarsama yöntemini bir araya getirerek oluşturulmuş bir yapıdır. Dolayısı ile sınıflandırma yapılmak istenmeyip sadece bir durumu eldeki özniteliklerden hangisi daha fazla etkiliyor ya da, eleme yapılmak istenilen durumlarda, hangi öznitelik benim işime yaramıyor gibi soruların cevabının arandığı durumlarda da kullanılabilme imkanı tanınmış olur (Çetişli, 2006).

## BULGULAR VE DEĞERLENDİRME

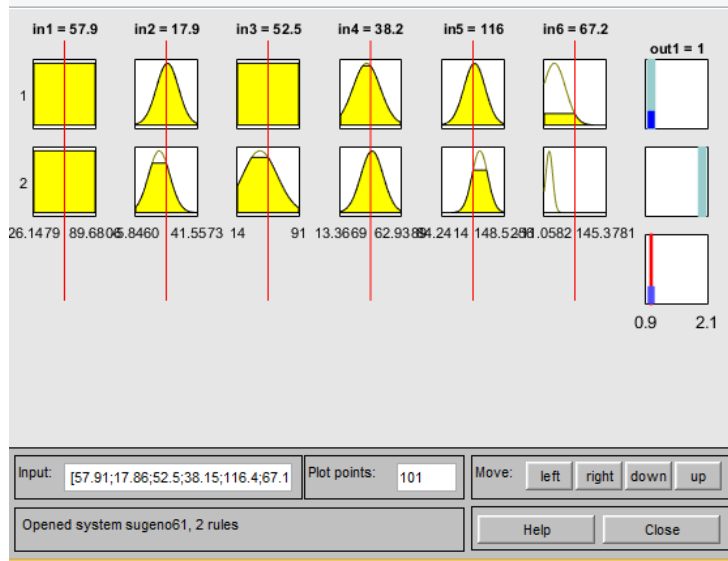
Bu çalışmada omurga durumu normal ve anormal olan toplamda 310 birey arasından hangilerinin normal omurgaya, hangilerinin ise anormal omurgaya sahip olduğu tespit edilmiştir. Bu sınıflandırma işleminin daha başarılı gerçekleştirilebilmesi için sınıflandırılmak istenen omurga durumunu etkileyen öznitelikler arasından önem derecesi daha yüksek olan öznitelikler tespit edilmiş, sınıflandırmada etkisinin olmadığı tespit edilen öznitelikler çıkarılmıştır. Böylece sınıflandırma işlemi gerçekleştirilirken önemsiz olduğu tespit edilen öznitelikler devre dışı bırakılmıştır. Sınıflandırma ve öznitelik seçiminde literatürde kabul görmüş birçok yöntem denenmiştir. Ulaşılan başarı değerlerinin her bir yöntemde farklılık gösterdiği gözlemlenmiştir. Hatta aynı yöntem içerisinde, test verileri ve eğitim verilerinin kullanılma oranı değiştirildiğinde bile farklı başarı oranlarının olduğu görülmüştür. Veri seti içerisinde önemli özniteliklerin tespit edilmesi ve etkisi olmayan özniteliklerin devre dışı bırakılmasında USBS yönteminden faydalanılmıştır. Etkisi olmayan öznitelikler tespit edilip çıkarıldıktan sonra sadece önemli görülen öznitelikler kullanılarak yapılan sınıflandırmada, test verisinde başarıya ulaştığı görülmüştür. USBS ile yapılan eğitimler 1000 iterasyon üzerinden gerçekleştirilmiştir. Veri setinin rasgele olarak %50 eğitim %50 test verisi olarak kullanıldığı zaman test verisinde %96.1935 oranında başarı elde edilmiştir. Bu sonuçlar elde edilirken öznitelikler için belirlenen önem derecesi Şekil 3 de, USBS'e ait bulanık kuralların modeli de Şekil 4 de gösterilmiştir.

Şekil 3 : Öznitelik Seçim Kriteri



Şekil 3'de de görüldüğü gibi önem derecesi az olarak nitelendirilen öznitelik lumbar lordosis angle değeri olurken, önem dereceleri yüksek olan öznitelikler pelvic tilt, sacral slope, pelvic radius ve degree spondylolisthesis olarak tespit edilmiştir. Pelvic incidence değerinin ise sonuca bir etkisini olmadığı görülmüştür. Pelvic incidence değerinin kümeleri ayrıştırmada bir etkiye sahip olmadığı tespit edildiği için sınıflandırma yapılırken devre dışı bırakılmıştır.

Şekil 4 : USBS'e ait bulanık kuralların modeli



USBS'nin ifade ettiği önemli özniteliklerin katkı dereceleri Şekil 4'de görülmektedir. USBS sınıflayıcısı; Tablo 1'de yer alan 2., 4., 5. ve 6. öznitelikler olan Pelvic tilt, Sacral slope, Pelvic Radius ve degree spondylolisthesis özniteliklerinin önemini yüksek derecede vurgulamıştır. Üçüncü öznitelik olan Lumbar lordosis angle özniteliklerinin sonuca katkısının az olduğunu birinci öznitelik olan Pelvic incidence özniteliklerinin de sonuca katkısının olmadığını ortaya konulmuştur.

#### KARŞILAŞTIRMA VE SONUÇ

Bu çalışmada insan omurgası veri seti üzerinde öznitelik seçimi gerçekleştirilmiştir. Böylece gerekli görülmeyen öznitelikler dışarda bırakılarak daha az öznitelikle daha doğru bir tahminleme gerçekleştirilmiştir. Yöntem olarak uyarlamalı sinir-bulanık sınıflayıcı (USBS) yöntemi kullanılmıştır. Kullanılan veri seti daha öncede kullanılmış ve çeşitli yöntemle analiz edilmiştir. Bu çalışmada ise üstünde durulan konu önemli özniteliklerin tespiti olmuştur. Bu yüzden dolayı literatürde öznitelik seçiminde sıkça kullanılan USBS yöntemi tercih edilmiştir. Çünkü sonucu etkileme açısından bakıldığı zaman, sınıflandırma içindeki her bir öz niteliğin önem derecesi aynı olmayabilir. Dolayısı ile doğru öznitelik seçimi ile başarı oranları artırılabilir. Bu çalışmada önem derecesi yüksek olan öz nitelikler USBS yöntemi ile tespit edilmiştir. Tespit edilen öznitelikler Pelvic Tilt, Sacral Slope, Pelvic Radius ve Degree Spondylolisthesis olmuştur. Önemli öznitelikler tespit edilmeden önce sınıflandırma başarı oranı %83 iken USBS Algoritması ile tespit edilen öznitelikler kullanıldığında ve yine USBS ile önemsiz olarak tespit edilen öznitelik sınıflandırma yapılırken devre dışı bırakıldığında, sınıflandırma başarısında kayda değer bir artış sağlanmıştır. Başarı performansı test verisi için %96.1935 şeklinde olmuştur. Böylece sınıflandırmada kullanılan USBS yöntemi ile öznitelik sayısı altıdan beşe düşürülmüş ve daha az öznitelik ile daha doğru bir sınıflandırma yapılmıştır. Öznitelik sayısının fazla olduğu durumlarda, başarı oranını düşürmeden sonuca etki değeri olmayan özniteliklerin tespit edilmesi ve

çıkarması günümüzde hızla artan veriler içerisinde anlamlı bilgiler çıkarılması açısından oldukça önemlidir. Bu süreç doğada bulunan madenlerin kullanılabilir duruma getirilmesi gibidir. Nasıl ki bütünden parçaya gidilerek var olan değerli kısımlara ulaşılırken rezerve sahip alanların doğru tespit edilmesi, maliyeti düşürmesinin yanında hızı artırıyor ise. Bu durum yapılan madencilik veri içerisinde olması durumunda da aynıdır. Böylece daha değersiz olan kısımlardan arındırılmış olan bilgi, elde edilirken maliyet azaltılmış ve verimlilik artırılmış olur.

Sonuç olarak; veri madenciliği tekniklerinde amaç var olan veriden işimize yarayacak olan bilgiyi elde etmektir. Elde edilecek başarı oranı ise veri ön işleme aşamasında ne kadar hassas çalışıldığı ile ilişkilidir. Özniteliklerin önem düzeyi ne kadar doğru tespit edilir ise elde bulunan veriye model oturtmak o kadar kolay olacaktır. Sonrasında oluşturulan bu modeller sayesinde daha doğru ve daha hızlı kararlar verilmesine imkan tanınmış olacaktır. Bu amaç doğrultusunda veri madenciliği teknikleri birçok alanda kullanılabilir. Özellikle tıp alanında insan sağlığı ile ilgili çalışmalarda kullanılabilir. Özellikle teşhis oranında her bir artış bile çok büyük önem derecelerine sahiptir. Bu yüzden birçok yöntem kullanılarak başarı değeri en yüksek olan sonuçlar kullanılmalıdır. Doğru öznitelik seçimi de başarı oranını artırdığı için, sadece sınıflandırma değil önemli öznitelik seçimleri içinde çeşitli yöntemler kullanılmalı ve önem derecesi fazla olan öznitelikler tespit edilmelidir. Böylece başarı oranı artırılırken, süreci hızlandırmak ve bellekten tasarruf etmek mümkün olacaktır. Aynı zamanda daha doğru yorumlamalarda bulunulmasına imkan tanınacaktır.

#### **KAYNAKÇA**

Akal, Ş. (2016). Sağlık alanında akıllı teknolojilerin kullanımı için bir uygulama önerisi. Uluslararası Yönetim Bilişim Sistemleri Konferansı, (3), 291-297.

Albayrak, A. S. ve Koltan Yılmaz, Ş. (2009). Veri madenciliği: karar ağacı algoritmaları ve İMKB verileri üzerine bir uygulama. Süleyman Demirel Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi, 14 (1), 31-52.

Altındağ, O. (2013). Kanser sınıflandırmada mikrona ve mrna anlatım bilgilerinin entegrasyonu. Yayınlanmamış yüksek lisans tezi, Başkent Üniversitesi, Ankara.

Altıntaş, Y. Y. (2010). Veri madenciliğinin tıpta kullanımı ve bir uygulama: hemodiyaliz hastaları için risk seviyelerine göre risk faktörlerinin etkileşimlerinin incelemesi. Yayınlanmamış yüksek lisans tezi, Gazi Üniversitesi, Ankara.

Berthonnaud, E., Dimnet, J., Roussouly, P., ve Labelle, H. (2005). Analysis of the sagittal balance of the spine and pelvis using shape and orientation parameters. *Clinical Spine Surgery*, 18(1), 40-47.

Budak, E. Ç., ve Bozkurt, M. R. (2013). Vertebra Lomber Disklerde Meydana Gelen Bozulmaların Manyetik Rezonans Görüntüleme (MRG) ile Analizi. *AJIT-e: Online Academic Journal of Information Technology*, 4(11).

Chae, Y. M., Ho, S. H., Cho, K. W., Lee, D. H., ve Ji, S. H. (2001). Data mining approach to policy analysis in a health insurance domain. *International journal of medical informatics*, 62(2-3), 103-111.

Coulter, D. M., Bate, A., Meyboom, R. H., Lindquist, M., ve Edwards, I. R. (2001). Antipsychotic drugs and heart muscle disorder in international pharmacovigilance: data mining study. *Bmj*, 322(7296), 1207-1209.

Çetinkaya Bozkurt, Ö. Ç., Kalkan, A. ve Çeşmeli, M. Ş. (2016). Karar destek sistemlerinin işletme yönetimi açısından önemi: mermer işletmelerinde bir araştırma. *Uluslararası Yönetim Bilişim Sistemleri Konferansı*, (3), 153-162.

Çetişli, B. (2006). Öznitelik seçiminde dilsel kuvvetli sinir bulanık sınıflayıcı kullanımı. *Eskişehir Osmangazi Üniversitesi Müh.Mim.Fak.Dergisi*, XIX (2), 109-130

da Rocha Neto, A. R., Sousa, R., Barreto, G. D. A., ve Cardoso, J. S. (2011, June). Diagnostic of pathology on the vertebral column with embedded reject option. In *Iberian Conference on Pattern Recognition and Image Analysis* (pp. 588-595). Springer, Berlin, Heidelberg.

da Rocha Neto, A. R., ve de Alencar Barreto, G. (2009). On the application of ensembles of classifiers to the diagnosis of pathologies of the vertebral column: A comparative analysis. *IEEE Latin America Transactions*, 7(4), 487-496.

Demircioğlu, H. Z. ve Bilge, H. Ş. (2015). Yumurtalık kanseri veri kümesindeki gen ifadelerinin veri madenciliği ile analizi. *Marmara Fen Bilimleri Dergisi*.

Fırat, m., Yurdusev, m. a., ve mermer, m. (2008). uyarlamalı sinirsel bulanık mantık yaklaşımı ile aylık su tüketiminin tahmini. *gazi üniversitesi mühendislik-mimarlık fakültesi dergisi*, 23(2).

Ganzert, S., Guttman, J., Kersting, K., Kuhlen, R., Putensen, C., Sydow, M., ve Kramer, S. (2002). Analysis of respiratory pressure-volume curves in intensive care medicine using inductive machine learning. *Artificial intelligence in medicine*, 26(1-2), 69-86.

Ghandhari, H., Hesarikia, H., Ameri, E., ve Noori, A. (2013). Assessment of normal sagittal alignment of the spine and pelvis in children and adolescents. *BioMed research international*, 2013.

Gharib, T. F., Nassar, H., Taha, M., ve Abraham, A. (2010). An efficient algorithm for incremental mining of temporal association rules. *Data ve Knowledge Engineering*, 69(8), 800-815.

Gökberk, B., İrfanoğlu, M. O., Doğu, H., Akarun, L., & Alpaydın, E. (2003). Yüz Tanıma İçin Eniyi Gabor Parametrelerinin Seçimi.

Göreke, V., Uzunhisarcıklı, E., ve Güven, A. (2014). Gri Seviyeli Eşoluşum Matrisleri Kullanılarak Sayısal Mamogram Görüntüsünden Doku Özniteliklerinin Çıkarılması ve Yapay Sinir Ağı ile Kitle Tespiti. *Tıp Teknolojileri Ulusal Kongresi-TıpTekno'14*.

Gündüz, A. E., Temizel, A., Temizel, T. T. (2013). Kalabalık Dinamiklerinin Çıkartılması İçin Özniteliklerin Tespit ve Takibi Feature Detection and Tracking for Extraction of Crowd Dynamics.

Hong, T. P., ve Wu, C. W. (2011). Mining rules from an incomplete dataset with a high missing rate. *Expert Systems with Applications*, 38(4), 3931-3936.

İşler, Y., ve Narin, A. (2012). WEKA Yazılımında k-Ortalama Algoritması Kullanılarak Konjestif Kalp Yetmezliği Hastalarının Teşhisi. *SDU Teknik Bilimler Dergisi*, 2(2).

Karasulu, B. (2016). Büyük veri çağında bilişim sistemleri için zeki teknikler destekli karar vermenin rolü: inceleme ve analiz. Uluslararası Yönetim Bilişim Sistemleri Konferansı, (3), 1-11.

Kızılkaya Aydoğan, E., Gencer, C. ve Akbulut, S. (2007). Veri madenciliği teknikleri ile bir kozmetik markanın ayrılan müşteri analizi ve müşteri bölümlenmesi. Mühendislik ve Fen Bilimleri Dergisi, 26 (1), 43-57.

Koyuncugil, A. S. ve Özgülbaş, N. (2009). Veri madenciliği: tıp ve sağlık hizmetlerinde kullanımı ve uygulamaları. Bilişim Teknolojileri Dergisi, 2 (2), 21-32.

Machine Learning Repository, <https://archive.ics.uci.edu/ml/datasets/Vertebral+Column>, (21.12.2016).

Onan, A., ve Korukoğlu, S. (2016). Metin sınıflandırmada öznelik seçim yöntemlerinin değerlendirilmesi. Akademik Bilişim.

Pençe, İ., Çetişli, B., Amasyalı, M. F., Çetin, M., Akbulut, C., Zihni, N. B., ve Kurban, R. (2013). El Yazı Karakterlerinin Kapalı Cebirsel Eğrilerle Modellenmesi Ve Sınıflandırılması. Sigma, 5, 1-7.

Pençe, İ. ve Tarhan, L. (2016). Öğrenim sürelerini uzatan üniversite öğrencilerinin veri madenciliği yöntemleriyle tahmini. Uluslararası Yönetim Bilişim Sistemleri Konferansı, (3), 291-297.

Reddy, S. K. & Kodali, S. R. & Gundabathina, J. L. (2012). Classification of vertebral column using naïve bayes technique. International Journal of Computer Applications, 58 (7), 38-42.