

## VERİ MADENCİLİĞİ İLE DEPREM VERİLERİNİN ANALİZİ

N. Duru<sup>-1</sup>, M. Canbay<sup>-1</sup>

**Posta Adresi:** <sup>1</sup>Kocaeli Üniversitesi Müh.Fak. Bilgisayar Mühendisliği

<sup>2</sup>Kocaeli Üniversitesi Müh.Fak. Jeofizik Mühendisliği

**E-posta:** nduru@kou.edu.tr, mucella@kou.edu.tr

---

**Anahtar Kelimeler:** Veri Madenciliği, Lineer Regresyon, Risk analizi

**ÖZ** Günümüzde bilgisayar sistemlerinin hızlı gelişmesine paralel olarak ilerleyen veri saklama sistemlerindeki gelişim, verilerin inanılmaz boyutlarda saklanmasına imkân sağlamaktadır. Sismik hareketler bu toplanan verilerden yalnızca birisidir. Bu çalışmada, YUBAM (Yeryüzü ve Uzay Bilimleri Araştırma Merkezi) dan alınan katalog veriler üzerinde veri madenciliğinde kullanılan tekniklerden biri olan lineer regresyon kullanılarak risk analizi yapılmaya çalışılmıştır. Risk analizi yapılacak bölgenin geçmişe yönelik deprem verilerinin alınması ve üzerinde veri madenciliği tekniği uygulamasının yapılabilmesi amacıyla bir yazılım geliştirilmiştir.

**Key Words:** Data mining, Linear regression, Risk analysis

**ABSTRACT** *Data mining or knowledge discovery in databases is the process of search for valuable information in large volumes of data, exploration and analysis of large quantities of data in order to discover meaningful patterns and relationships that may be used to make valid predictions. The prediction of the earthquakes is a very difficult and challenging task. In this study, we have developed a software to use linear regression analysis, which is a technique of data mining, for predicting earth quake.*

## GİRİŞ

Verilerin günlük artışının son 20 yılda inanılmaz boyutlara ulaşması ile bu verilerin uzmanlar tarafından incelenmesi imkansız bir hal almaya başlamıştır. Bu nedenle yeni yöntemler aranmaya başlamıştır. Bu yöntemlerden en gözde olanı veri madenciliğidir. Ancak veriyi anlamlı hale getirmek için geçen sürece veri analizi denir. Veri yığınının anlamlı bilgi elde etmek için çeşitli analizlerden geçirilir bu analizlerin bütününe veri madenciliği denir. Veri madenciliği büyük miktarda veri içinden gelecekle ilgili tahmin yapmamızı sağlayacak bağıntı ve kuralların bilgisayar programları kullanarak aranmasıdır gibi bir tanımlamada bulunulabilir. Günümüzde, bir marketten yapılan alışverişin fiş hareketi bile tutulmaktadır. Büyüme işlevleri cinsinden ifade edecek olursak, veri saklama kapasitesi her 9 ayda bir tahmini ikiye katlanmakta (Porter,J., 1998), buna karşılık aynı periyotta, Moore kanununa göre hesaplama gücü iki kat daha az büyümektedir Braynt, R.E ve diğ., 2003 Verideki bu patlama sonucunda, büyük veritabanlarından değerli, ilginç ve önceden bilinmeyen bilgiyi keşfetmek için pratik uygulamalar ve olası çözümler için önemli ve aktif bir araştırma alanı olan veri tabanlarında bilgi keşfi (VTBK) ortaya çıkmıştır (Altıntop Ü., 2006). Veri madenciliği ve VTBK sık sık, eşanlı olarak kullanılırken, veri madenciliği, gerçekte bilgi keşif sürecinin bir parçasıdır.

Veri madenciliğinde kullanılan modeller, tahmin edici (Predictive) ve tanımlayıcı (Descriptive) olmak üzere iki ana başlık altında incelenmektedir.

•Tahmin edici modellerde, sonuçları bilinen verilerden hareket edilerek bir model geliştirilmesi ve kurulan bu modelden yararlanılarak sonuçlan bilinmeyen veri kümeleri için sonuç değerlerin tahmin edilmesi amaçlanmaktadır.

•Tanımlayıcı modellerde ise karar vermeye rehberlik etmede kullanılabilir mevcut verilerdeki örüntülerin tanımlanması sağlanmaktadır.

Veri madenciliği modellerini gördükleri işlemlere göre,

1- Sınıflama (Classification) ve Regresyon (Regression)

2- Kümeleme (Clustering)

3- Birliktelik Kuralları (Association Rules)

olmak üzere üç ana başlık altında incelemek mümkündür. Sınıflama ve regresyon modelleri tahmin edici, kümeleme ve birliktelik kuralları modelleri tanımlayıcı modellerdir.

Veri madenciliği her geçen gün yeni alanlarda uygulanmaktadır. Özmen,Ş.,2003 kredi kartı uygulamasında istatistik ve veri madenciliği tekniklerinden yararlanarak kimlere kredi kartı verilebileceği limit ve kartın cinsi gibi hususları ele almıştır. Mevcut müşteriler ve potansiyel müşteriler yıllık gelir, borç ödeme süreleri, geciktirme sayıları vb koşullara göre sınıflandırılmıştır. Sınıflandırma sonucunda kredi miktarları ve verilip verilmeyeceği ortaya çıkmıştır.

Han ve diğ.,2001 çalışmalarında mezun olan öğrencilerin çalıştıkları sektördeki deneyim-maaş ilişkisini incelemişlerdir. Çalışmalarında lineer regresyon kullanarak yıllara göre maaş tahmini yapmaya çalışmışlardır.

Bu çalışma, Kilikya bölgesine ait deprem verilerinin 4 şiddet ve üstünü kapsamakta olup 1900 yıllık bir süreçte bu verilerin yıllara göre dağılımlarında boşluklar bulunduğu göze çarpmaktadır.Bu durum veri sayısını azaltmakta yanıltıcı sonuçlara sebep olabilmektedir.Bu gibi koşullarda veri madenciliği problemlerinden birinin çözümüne yönelik, bilgi keşfi modeli belirli bir güvenlik derecesinde tahmini kararlar alabilmektedir.

### **Metot**

Bu çalışmada veri madenciliği tekniklerinden lineer regresyon analizi uygulanarak deprem verileri analiz edilerek risk tahmini yapılmaya çalışılmıştır. Uygulamada ilk olarak excelde tutulan katalog verilerinin alınması için Excel okuma programı tasarlanmıştır. İlgili formüllerin analiz edilmesi ile program geliştirilmiş ve risk analizi yapacak hale getirilmiştir. Excelden elde edilen veriler veri tabanına kaydedilmeden önce gereksiz yinelenmeler ve null değerlerden arındırılmıştır. Uygulama MS Access veri tabanı ve Microsoft Visual Studio kullanılarak gerçekleştirilmiştir.

Lineer regresyon analizinin amacı iki değişken arasındaki gerçek ilişkinin (1) numaralı denklem yardımıyla bir tahminini elde etmektir. Diğer bir deyimle, değişkenler arasında bulunduğu kabul edilen gerçek doğrusal ilişkiyi,

$$Y=a +\beta X$$

(1)

ile gösterilirse, regresyon sonucunda ortaya çıkacak olan  $a$  ve  $b$  katsayıları gerçek  $a$  ve  $\beta$  parametrelerinin birer tahmini mahiyetinde olacaktır. Bu varsayımın önemli avantajı, ileride de görüleceği gibi, gerçek parametreler ile ilgili güven aralıkları hesaplamayı ve hipotez testleri yapmayı mümkün kılmasıdır (Köksal,B,A.,2003). Şekil.1 de, giriş ekranında "Yeni Veri Ekle" tıkladığında ekrana gelen ara yüz örneği verilmektedir.

Yeni Veri Ekle

Yeni Kayıt

Yıl.....:  Enlem.....:

Ay.....:  Boylam.....:

Gün.....:  Derinlik.....:

Saat.....:  Magnitud.....:

Dakika.....:

Saniye.....:

Temizle Kaydet

En Son Kayıt

ID	Yıl	Ay	Gun	Saat	Dk	Sn	Enlem	Boylam	Derinli	Magni
60109	2004	12	31	18	47	45	37,83	30,25	5,000	28,68
*										

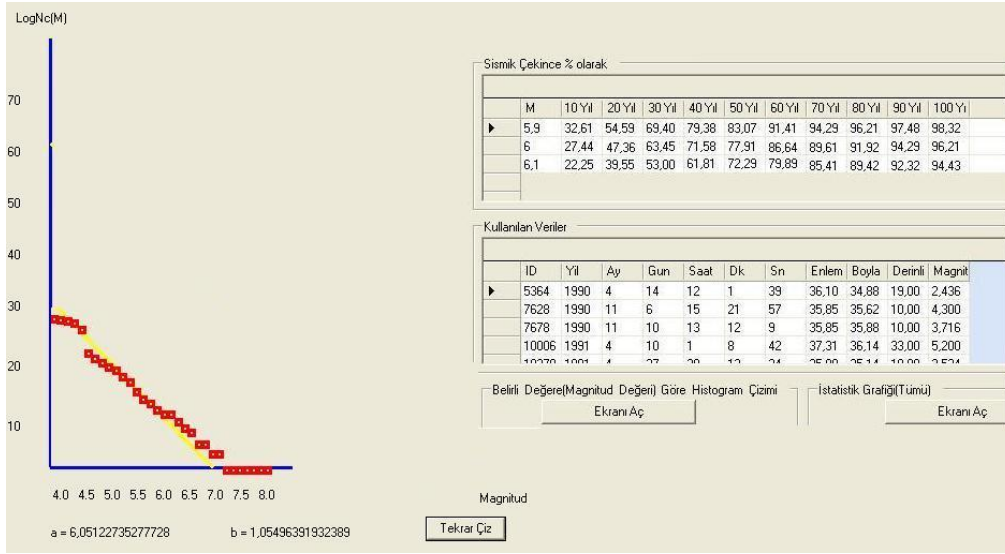
Excelden Veri Alma

Excelden Verileri Al Yardım

Şekil-1. Yeni Veri Ekle ara yüzü

Bu ara yüz iki kısımdan oluşmaktadır: Yeni Kayıt ve Excelden Veri Alma.

Yeni Kayıt kısmında veriler tek tek girilmek durumundadır. Toplu eklemeler için kesinlikle tavsiye edilmez. Excelden Veri Alma kısmında ise veriler toplu olarak istenilen Excel dosyasından yapılabilmektedir. Veriler veri tabanında tutulduğundan dolayı veri ekleme bazı standart işlemleri beraberinde getirmektedir. Veriler MS Access veritabanındaki Deprem adlı tabloda tutulmaktadır. Tabloda sırasıyla Yıl, Ay, Gun, Saat, Dk(dakika), Sn(saniye), Enlem, Boylam, Derinlik, Magnitüd isimli sütunlar bulunmaktadır. Veri ekleme yapılırken bu sıranın göz önünde bulundurulması gerekmektedir.



Şekil-2. Sismik Tehlike Verileri

Şekil-2'de görülen arayüzde kullanıcı, özellikle istenilen bir magnitüd değerine göre veya verilen Regresyon Eğrisi Aralık Değerine göre depremlerin risk analizini yapabilmektedir.

Kullanıcı belirli bir bölgenin enlem, boylam ve bu bölge kare olacağından enlem ve boylama ne kadarlık bir ekleme yapılacaksa o ekleme miktarını girer. Örneğin Çukurova ve çevresi için 35,5–38,0 enlem ve 34,5–37,0 boylam değerini kullanmaktadır. Programa girilecek değer 35,5 enlem, 34,5 boylam ve 2,5 aralık değeridir.

## TARTIŞMA VE SONUÇ

Bu çalışma, deprem verileri kullanılarak seçilen bir bölgeye ait sismik tehlikenin diğer deyişle gerçekleşme olasılığının veri madenciliği yönünden ele alınarak incelenmesini kapsamaktadır. Çalışma sonuçları Jeofizik sonuçlar ile korele edilerek doğruluk payı da araştırılmıştır (Ulutaş,E.,1999). Her gelecek 10 yıl için % sismik tehlike değeri artış göstererek devam etmiş, örneğin 6 magnitüdüdeki bir depremin olma olasılığı 10 yıl içinde %27 iken, 30 yıl içinde %60 ve 60 yıl için de %80 leri bulmaktadır. Bu değerler daha önce çalışma bölgesinde yapılmış çalışmalarla uyum göstermektedir. Ancak burada unutulmaması gereken bu çalışmanın deprem tahmini için kullanılan tekniklerden sadece birisi olduğu ve bu çalışmanın konusu itibariyle çalışma bölgelerinin tektonik özelliklerini hiç irdelemeden dahi olsa olumlu sonuçlara varılabilmemesinin mümkün olduğunun gösterilebilmesidir. Bu durum da ayrıca çalışma sonuçları açısından sevindirici olmuştur. Ayrıca yapılan çalışmanın sonuçlarının büyük bölgelere göre küçük bölgelerde daha iyi sonuç verdiği görülmesidir. Uygulama, dünya ölçeğindeki her noktanın analizini yapacak şekilde geliştirilmiş olup, ihtiyaç halinde programa eklemeler yapmak suretiyle başka bu tür çalışmalar yapacak şekilde tasarlanmıştır. Sonuç olarak, bu çalışma türünden bazı istatistik yöntemlerin de kullanılabilirliği ve faydalı olabileceği ve veri miktarının önemli bir kriter olduğu, ne kadar çok veri üzerinde çalışılırsa o derece başarılı sonuçlar alınmasının mümkün olacağı söylenebilir.

## KATKI BELİRTME VE TEŞEKKÜR

Deprem kataloglarından veri seçiminde ve bölgelendirilme çalışmasında bilgi ve yorumlarını gördüğümüz Kocaeli Üniversitesi YUBAM (Yeryüzü ve Uzay Bilimleri Araştırma Merkezi) öğretim üye ve elemanlarına, Ayrıca içtenlikle tecrübelerini ve çalışmasını bizimle paylaşan Arş.Gör.Dr. Ergin Ulutaş'a teşekkür ederiz.

## KAYNAKLAR

Altıntop Ü., 2006. İnternet Tabanlı Öğretimde Veri Madenciliği Tekniklerinin Uygulanması. **Yüksek Lisans Tezi**. Kocaeli Üniversitesi Fen Bilimleri Enstitüsü

Braynt,R.E. ve O'Hallaron, D.R.. 2003. **Computer Systems: A Programmer's Perspective**, Prentice Hall, New Jersey, ABD,127

Han, J., Kamber, M., 2001. **Data Mining: Concepts and Techniques**, Morgan Kaufmann Publishers, San Francisco.

Köksal, B.A., 2003. **İstatistik Analiz Metodları**. Çağlayan Kitabevi. İstanbul

Özmen Ş., 2003. İş Hayatında Veri Madenciliği Yeniden Keşfediliyor. Marmara Üniversitesi I.I.B.F. Yayınları

Porter, J. 1998. Disk Trend 1998 Report. [www.disktrend.com/pdf/portrpkg.pdf](http://www.disktrend.com/pdf/portrpkg.pdf)

Ulutaş,E.,1999.\_Çukurova ve çevresinin deprem tehlikesi, **Yüksek lisans tezi**,\_Kocaeli Üniversitesi Fen Bilimleri Enstitüsü