

VERI MADENCILIĞİ

Fırat İsmailoğlu, PhD

Giriș – Motivasyon - Genel Bakış



Ders Sorumlusu:

Ar. Gör. Dr. Fırat İsmailoğlu

Bilgisayar Mühendisliği, Oda No:202

fismailoglu@cumhuriyet.edu.tr

Ders Hakkında:

- 12 Hafta Ders
- Kitap I: Introduction to Data Mining, Pang-Ning Tan, Micheal Steinbach, Vipin Kumar
- Kitap 2: Data Mining the Textbook, Charu Aggarwal
- Yaz okulu olmayacak
- Değerlendirme: %40 Vize + %60 Final
- Belki Ödev/Proje?



Veri Madenciliği Nedir?

Veri Madenciliği (VM) (büyük) veri depolarından işe yarar, ilginç ve gizli bilgileri çıkarma sürecine denir.





VM ile ortaya çıkarılan bilgiler kullanılarak, elimizdeki veriyi daha iyi taniyabilir (tanımlayici analiz) ve gelecek hakkında tahminde bulunabiliriz (öngörücü analiz).



Eldeki veriyi kullanarak, verinin icindeki modeli/motifi (pattern) aramak, gelecek veriler hakkinda tahminde bulunmak aslinda yeni bir uğraş degildir:

ör. Eski Mısır'da Mısırlıların Nil nehrinin daha önceki taşma dönemlerini inceleyerek Nil'in gelecek defa ne zaman taşacağını tahmin etmesi.

ör. 6,11,21,41,81 sayı dizisi verilsin. Bu dizi (bu veri) incelenirse birinci sayi 6 olmak uzere her sayi bir oncekinin 2 katından bir eksiktir ($a_1 = 6$; $a_n = 2a_{n-1} - 1$). Bu model elimizdeki veriyi acıklar. Hatta bir sonraki sayiyi tahmin edebilririz: $2 \cdot 81 - 1 = 161$.

Veri madenciligi ise aynı sekilde eldeki verileri inceler, bu verileri açıklayıcı bır model geliştirir ve bazen gelecek veriler için tahminde bulunur.

Fakat veri madenciliğinde veri her sev olabilir, her verden gelebilir.





- İnternet Verileri: Kullanicilarin websitelerine erişimi sonrasi oluşan 'web access log' dosyaları. Bu dosyalar serverlarda saklanır. Bu dosyalarda kullaniciların IP adresleri, websitesine erişim zamanı, toplam erişim suresi ve ziyaret edilen sayfalar gibi bilgiler bulunur. Bu dosyalar üzerinde madencilik yapılarak belirli bir websitesinde geçirilen ortalama zaman, websitesine hangi bölgelerden daha çok erişim olduğu ve dığger kullanıcı aliskanlıkları bulunabilir.
- Finans Verileri: Kredi karti islemleri, kredi başvurusu dosyalari. Bu alanda veri madenciligi yapilarak olagan disi işlemler bulunabilir, çalinti kart kullanımi yakalabilir. Yada bankaya kredi icin basvuran kisinin alacagi krediyi ödeyip odememe durumu tahmin edilebilir.
- Alışveriş Verileri: Bir süpermakette gün boyu yapilan alışverişlerin olusturdugu veriler uzerine veri madenciligi yapilarak hangi ürünlerin beraber satin alindigi; yada müşteri profilleri (yaş, cinsiyet, gelir) ile aldıkları urunler arasında iliskiler bulunabilir. Kişiye özel kampanya yapılabilir.



- Sensör Verileri: Günümüzde sensör teknolojisinin yayginlasmasi ve ucuzlamasi ile saatten, çamaşır makinesine birçok ürüne yerleştirilen sensörlerin oluşturduğu gerçek zamanli veriler. Bu veriler uzerine yapilan veri madenciligi calismalariyla ornegin uzerinde sensör olan ürünün ne zaman bozulabilecegi ile ilgili tahmin yapilabilinir.
- Kullanici Verileri: Örneğin cep telefonu ile yaptigimiz konusmalar, ne sıkılıkla konustugumuz, kimleri aradigimiz, ne kadar uzun konustugumuzun verileri operatorlerde toplanir. Bu veriler uzerine tyapilan veri madenciligi calismalariyla kisinin konusma aliskanlıkları ortaya cikarilir ve boylece kisisel kampanyalar olusturabiliriz.

