

Veri Bilimi

Engin Demir, PhD.

Çankaya Üniversitesi

Bilgisayar Mühendisliği Bölümü

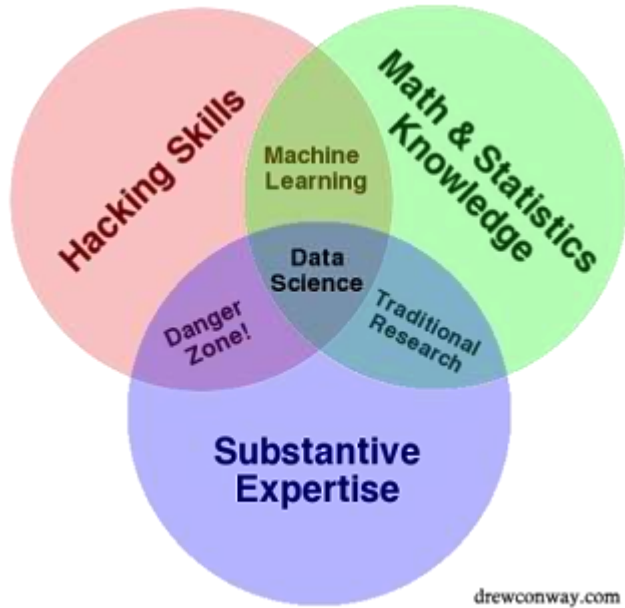


Veri Bilimi Nedir

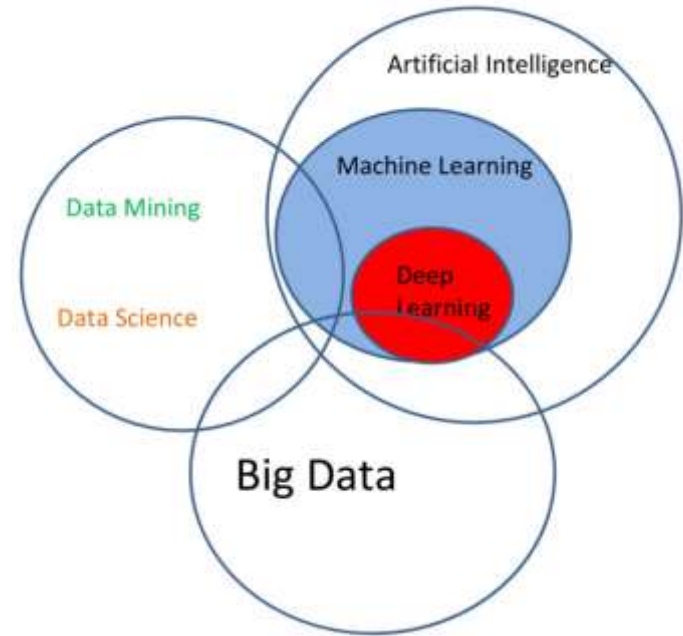
- Temelde çok büyük verinin otomatik yöntemlerle analiz edilmesi ve bilginin çıkartılmasıdır.
- Çok disiplinli bir araştırma alanıdır: Bilgisayar Bilimleri, Modelleme, İstatistik, Analitik ve Matematik
- Özellikle son yıllarda yeni bilim dallarının olumasında etkisi büyüktür: genetikten yüksek enerji fizikine
- İnsani ve sosyal bilimlere etkisi büyüktür
- Endüstride sağlıktan medyaya birçok alanda değişimin öncüsüdür



Veri Bilimi



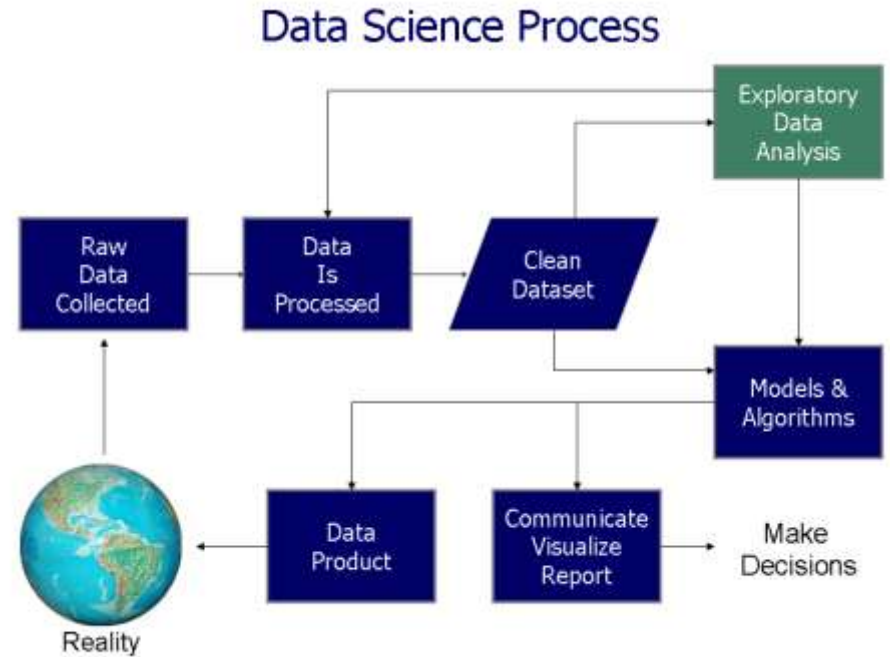
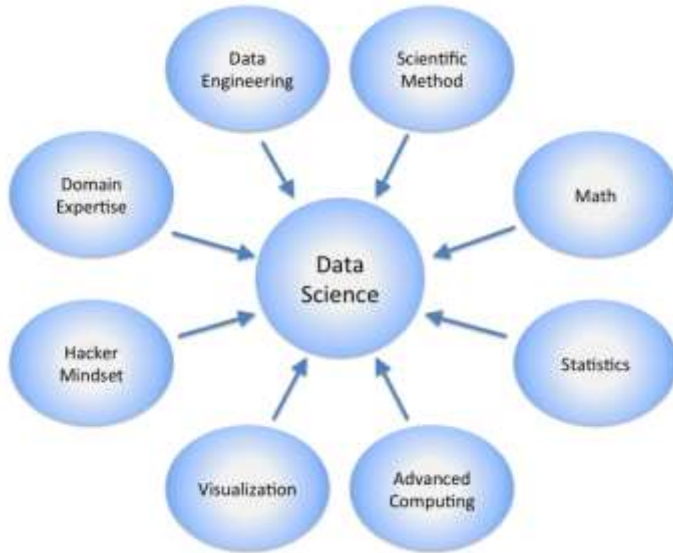
[Drew Conway in 2010](#)



Matthew Mayo [blogged a diagram](#) by Gregory Piatetsky-Shapiro in 2016



Yetenekler ve Süreç



Wikipedia: Data Science



Geneksel Yöntemler

- Veri toplama:
 - Algılayıcılar
- Veri üretme:
 - Deneyler
 - Gözlemler
 - Simülasyonlar
- Veri yönetimi
 - Verinin saklanması
 - İndeklenmesi
 - Sorgulanması ve getirilmesi
- Veriden bilginin türetilmesi
 - Veri madenciliği: istatistik ve makine öğrenmesi



Büyük Veri

- Hacim (Volume)
Büyük miktarda: Zettabytes($\sim 10^{21}$)
- Hız (Velocity)
Sürekli büyümekte: duran ve akan veri
- Çeşitlilik (Variety)
Çeşitli yapıda: metin , görüntü, ses, video
- Doğruluk (Veracity)
Kalite ve doğruluk kontrolü
- Değer (Value)
Kıymetli bilgi ve ürün üretmek



Büyük Veri Yönetimi ve Analizinde Zorluklar

Committee on the Analysis of Massive Data, Committee on Applied and Theoretical Statistics, Board on Mathematical Sciences and Their Applications, Division on Engineering and Physical Sciences, National Research Council. *Frontiers in Massive Data Analysis*. The National Academic Press. 2013.

- Çok dağıtık veri kaynakları ile ilgilenmek,
- Veri kaynağı izleme: veri üretiminden veri hazırlamaya,
- Veriyi geçerleme,
- Örnekleme önyargıları ve çok çeşitlilik ile başa çıkmak,
- Farklı veri formatları ve yapıları ile çalışmak,
- Paralel ve dağıtık mimarilerden faydalanan algoritmalar geliştirmek,



Büyük Veri Yönetimi ve Analizinde Zorluklar

- Veri bütünlüğünü sağlamak,
- Veri güvenliğini sağlamak,
- Veri keşif ve bütünleştirmeyi etkinleştirmek,
- Veri paylaşımını etkinleştirmek,
- Büyük veriyi görselleştirmek için yöntemler geliştirmek,
- Ölçeklenebilir ve artan algoritmalar geliştirmek,
- Gerçek zamanlı analiz ve karar verme ihtiyacı ile başa çıkabilmek

