

Veri Bilimi Nedir?

Üzerinde yaşadığımız dijital çağda, internet üzerinde üretilen aktivitelerin bir sonucu olarak devasa miktarlarda yapısal ve yapısal olmayan veri yığınları oluşmuştur. İnterneti yaygın olarak kullanmaya başladığınız yılları düşünün... 2003 yılında insanlık bir yılda, tüm insanlık tarihi boyunca üretilen veri kadar veri üretti ve aynı hızda üretmeye devam ediyor. Öyle ki IDC istatistiklerine göre 2020'de ulaşılacak veri miktarı, 2009'un 44 katı olacak.

Çok yakında dünyadaki tüm verilerin yaklaşık yarısının online ve birbiri ile bağlantılı verilerden oluşacağı tahmin edilmekte.* Dünya önümüzdeki günlerde veriyle gelişen teknolojiler nedeniyle bambaşka bir dünya olacak. Tüm bu gelişmeler için üretilen teknolojilere ve bu veri yığına kısaca **Big Data** yani **Büyük Veri** deniliyor.

Büyük Veri, doğru analiz araçları ile kullanıldığında işletmelerin stratejik kararlarını almalarında ve risklerini yönetmelerinde önemli bir rol oynuyor. Öncü işletmeler; yapısal olmayan verilerden oluşan bu **Büyük Veri**'yi işleyip anlamlı aksiyon planları oluşturarak, **Veri'den Para'ya** ** yaşam döngüleri oluşturuyorlar.

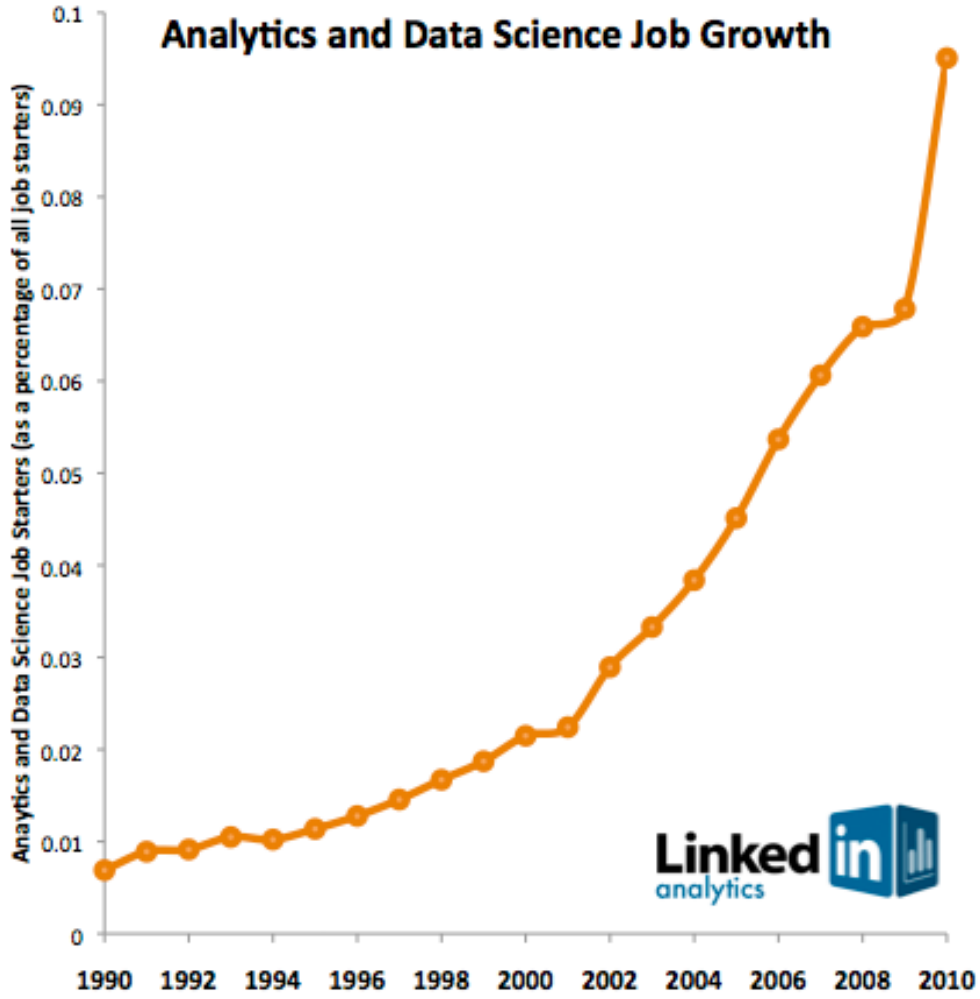
Büyük Veri'nin artış hızı, bu analizlerin yapılması için gerekli olan iş gücü ihtiyacını da hızla artırdı. Bu durum; eldeki veriyi yapılandıran, veri modelleri üreten ve bu modelleri ürün ya da hizmet süreçlerinde ve karar mekanizmalarında kullanabilen çok değerli bir alan olan **Veri Bilimi**'ni ortaya çıkardı.

Veri Bilimi alanının son yılların gözde mesleklerinden biri olduğunu alttaki grafiklere bakarak da görebiliriz:

Veri Bilimi Google Arama Trendleri

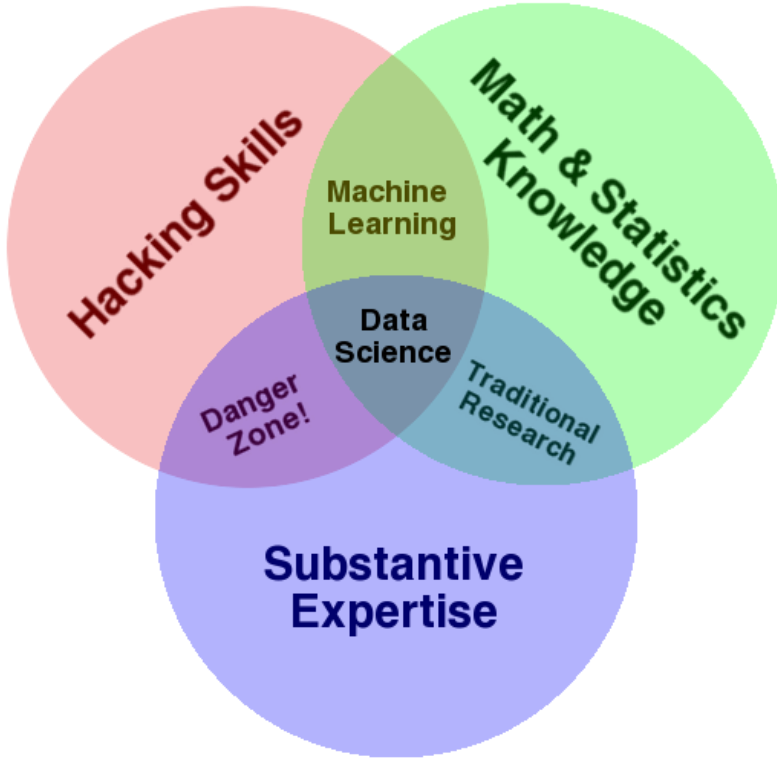
Veri Bilimi İş İlanları

Veri Analitiği iş ilanlarında da son yıllarda kayda değer bir artış gerçekleşmiş.



Veri Bilimi Alanını Tanıyalım

Veri Bilimi çok multidisipliner bir meslektir ve bu alanda çok sayıda alandan bir çok teori ve teknik kullanılır. Bunlar kısaca; sinyal işleme, matematik, olasılık modelleri, makine öğrenmesi, istatistik, bilgisayar programlama, veri mühendisliği, görselleştirme, veri ambarı gibi alanlardır. Veri Bilimi konusunda disiplinler arası ilişkileri gösteren **en çok kullanılan venn diyagramı alttaki venn diyagramıdır:** ***

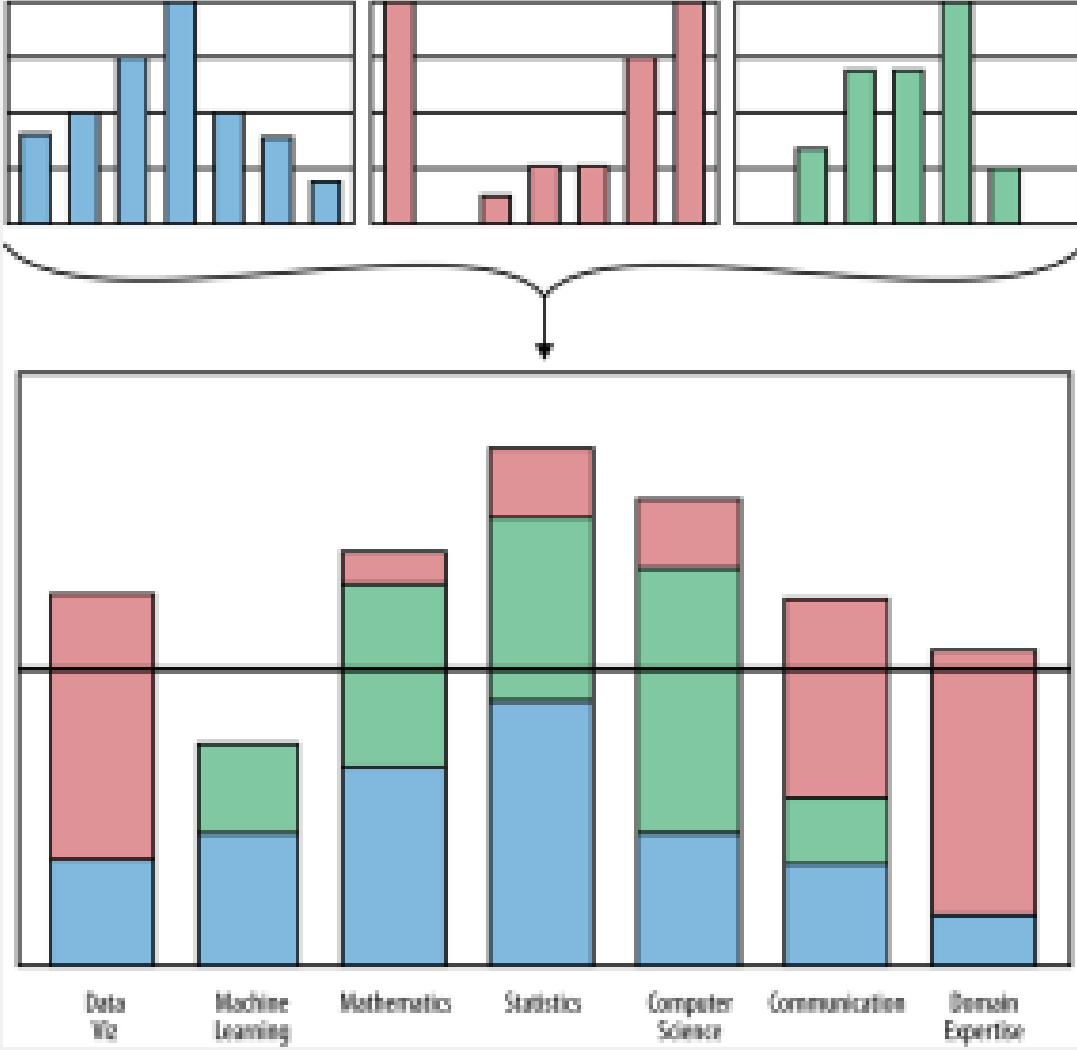


Veriden anlam çıkaran bu multidisipliner alanda Veri Bilimcileri farklı kılan bir takım özellikler bulunmaktadır.

Bu alanda öne çıkan konular ise resimde detaylı bir şekilde gösterilmiştir:

Fakat her **Veri Bilimci**'nin tüm bu özelliklere aynı anda sahip olmasını bekleyemezsiniz. Bu nedenle işletmelerde özellikleri ile bir birlerini tamamlayan **Veri Bilimi** takımları oluşturulmaktadır. *****

No one person can be the perfect data scientist, so **we need teams.**

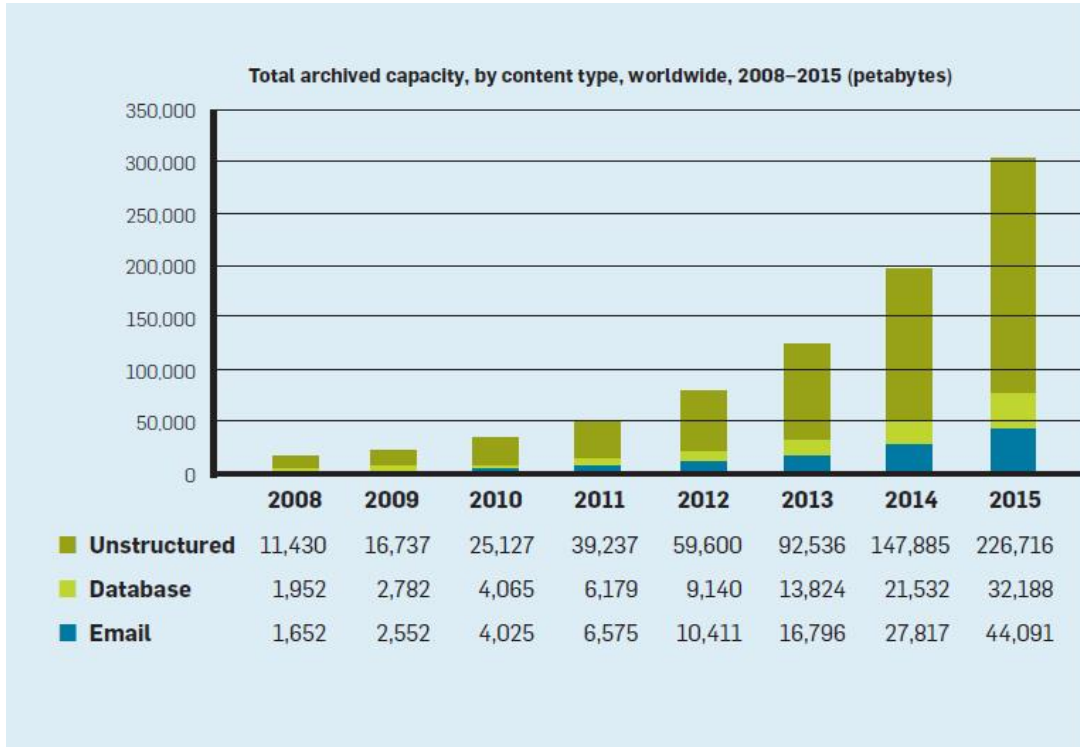


Veri Bilimi Takımı

Hangi özellikleri Veri Bilimi alanını bilim, Veri Bilimci'yi "Bilimci" olarak tanımlatıyor?

Peki istatistiğe çok benzeyen bir alanı neden yüzyıllardır kullanılan bir isim olan istatistik adında tanımlamayıp "Veri Bilimi" adıyla adlandırıyoruz. Veri Bilimi alanını istatistik biliminden ve diğer alanlardan ayıran bir kaç durum vardır:

– Birincisi bu işin hammaddesi olan, data çok heterojen bir yapıdadır. Ve unstructured(yapılandırılmamış) bir haldedir. (text, images, video)



Resimde 2008 yılından 2015 yılına kadar tahmini “unstructured” veri miktarını görüyorsunuz. İki farklı tipteki datayı analiz edip birleştirmek ve mantıklı bir hale getirmek bazen bilgisayar bilimi, dil bilimi, ekonomi, sosyoloji vb diğer alanları da konuşurmayı gerektirir. Mühendislik perspektifinden bakılınca bu denli verinin keşfedilmesi için geleneksel veritabanı sistemi yöntemleri yetersiz kalmaktadır. Geleneksel veri tabanı sistemleri bu tip işlemler için yetersizdir. Çünkü onlar veriye hızlı erişim ve hızlı veri özetlemek için optimize edilmişlerdir. Kullanıcı neyi sorarsa ona cevap verir. Çok iyi formüle edilmiş bir query bile Data’nın büyük arazisindeki desenleri ortaya çıkarmak için yetersiz kalırlar.

[John Tukey](#) bir konuşmasında şunu söylüyor: <Bazı dataların anlamlı kombinasyonu, büyük bir veri yığınının çok daha fazla işe yarayabilir. 100 GB’lık veriniz işe yaramaz, ama 3 KB’lık veri sorunuzun cevabını içerebilir.> Bu verinin multidisipliner bir şekilde seçilmesi, ayıklanması ve işlenmesi gerekmektedir.

Veri bilimi soru ile başlar. Bir soru sorduğunuzda yeni veri toplayıp toplamamanız gerektiğine de karar verirsiniz. Ve bu veriden doğru bir sonuç elde edip etmediğinizi deneysel bir şekilde değerlendirebilirsiniz. İşte bu nedenle bu alan bilimsel bir alandır. Veri Setinizde yapı ya da ağ bulmak kolaydır. Eğer yeteri miktarda veri toplarsanız, daima binlerce neden için korelasyonlar olacaktır. **Science** kelimesinin özellikle bu alanda kullanmasının nedeni, sistematik çalışmalarla bilgiye ulaşılması nedeniyledir. Bir başka tanım Test edilebilir açıklamalar ve tahminler içerisinde sistematik bir bilgi inşası ve bilgi organizasyonu olduğudur.