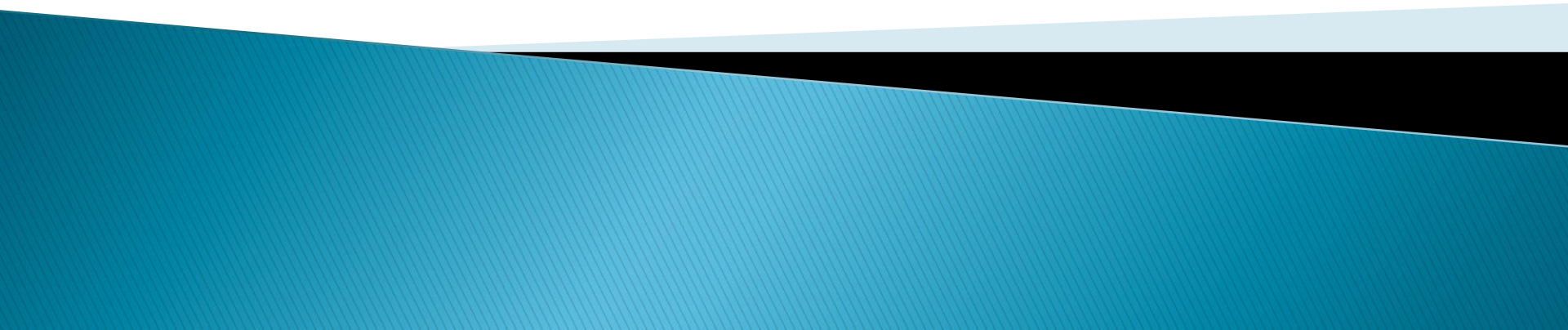


DEV VERİ VE VERİ MADECİLİĞİ

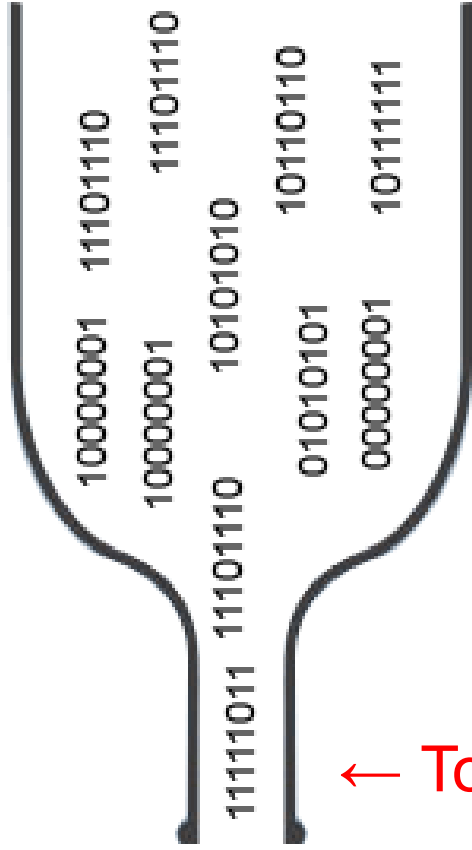
- **Herkes herşeyi, kaydediyor, saklıyor**
- **Internet kaydi, e-ticaret**
- **Her türlü alışveriş**
- **Bankalar**
- **Şirketler**
- **Hükümetler**
- **Sosyal ağlar**
- **Sosyal medya**

- **Internet kullanıcı sayısı 2014'te 3 milyar**
 - **Günde ortalama 2.5EB veri**
 - **1 Exabyte = 10^6 TB = 10^{18} Byte**
 - **Son iki yılda %90**
- 

- **Dünyadaki toplam bilişim harcamaları yılda ortalama %5 artarken, üretilen veri miktarı ise ortalama %40 artmaktadır.**
- **Bu inanılmaz miktardaki veriler hayatın hemen her alanından gelmekte olup, bunların toplanması, saklanması, güvenliği ve analizi bugün araştırmacılar için en önemli sorun haline gelmiştir.**

- **İnsan geni: 4GB**
 - **ABD Kongre Kütüphanesi: 250TB – 2012**
 - **NASA Meteoroloji Simulasyonu: 35PB – 2012**
 - **Facebook: 300PB + 500TB/per day – 2013**
 - **Google: 300PB/per day – 2013**
- 

- **Dev veri; toplumsal medya paylaşımları, ağ günlükleri, bloglar, fotoğraf, video, log dosyaları v.b. gibi değişik kaynaklardan toparlanan tüm ham kayıtlardan oluşan veridir.**
- **Bu verilerin hem çok büyük olmaları, hem farklı kaynaklardan toplanması, hem de belirli bir düzen içinde olmamaları sebebiyle standart yaklaşımlar, gerçek zamanlı analizlere imkan vermemekte**



← Toplanabilen veriler

← Toplanabilen verilerin incelenebilen kısmı

- Three 'V's:
- Volume (large amounts of data) - **BOYUTLARI**
- Velocity (rapidly changing data) - **DEĞİŞKENLİĞİ**
- Variety (different data sources and formats) – **ÇEŞİTLİLİĞİ**



SOSYAL MEDYA SOSYAL AĞ

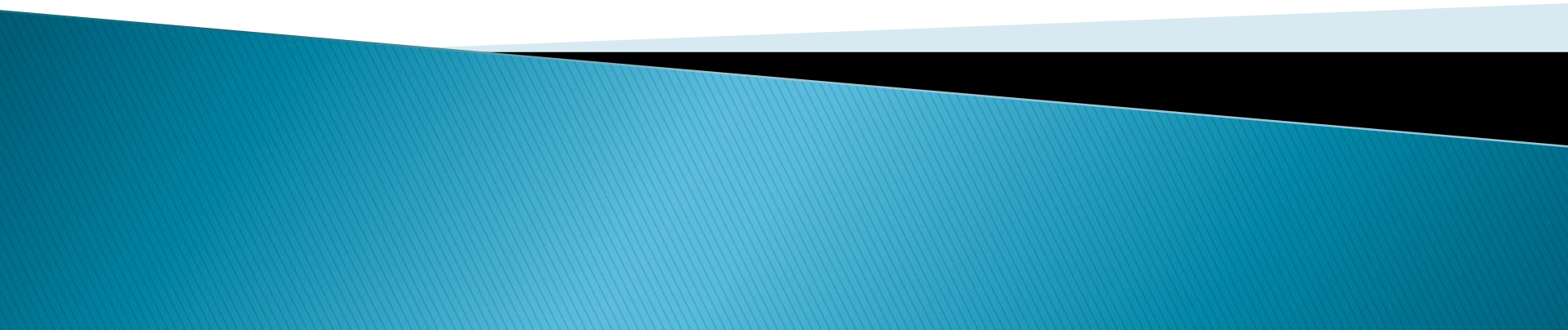


BİLİMSEL EKİPMANLAR



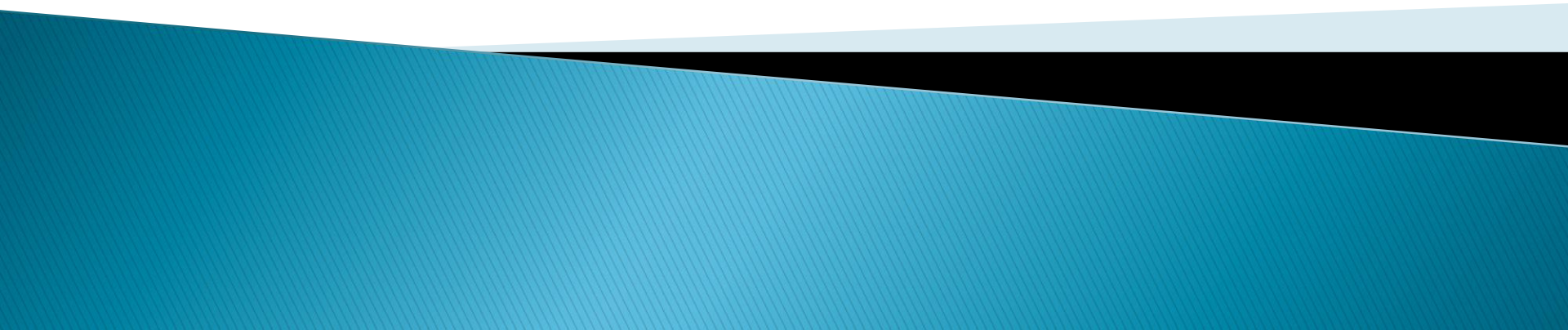
MOBİL CİHAZLAR

- Bornova Tansaş Merkez Şubede ne oluyor?
- Twitter'da Türkiye'de en popüler konular neler?
- Bayramda en çok hangi ürün satılıyor?

- **CRM: Müşteri Memnuniyeti**
 - **Pazar Araştırmaları**
 - **Performans arttırımı**
 - **Sistem optimizasyonu**
- 

□ Verilerle ne yapılabilir?

- Kümeleştirme
- Endeksleme, arama yapma, karşılaştırma
- Bilgi çıkarımı

- **Doğru veri nasıl bulunur?**
 - **Veri nasıl hızlı analiz edilebilir?**
 - **Veriden işe yarayan sonuç nasıl çıkartılır?**
- 

□ **Sonuca ulaşmak için:**

- **İstatistik**
- **Makine öğretisi**
- **Karşılaştırma**
- **Hepsinin karışımı**

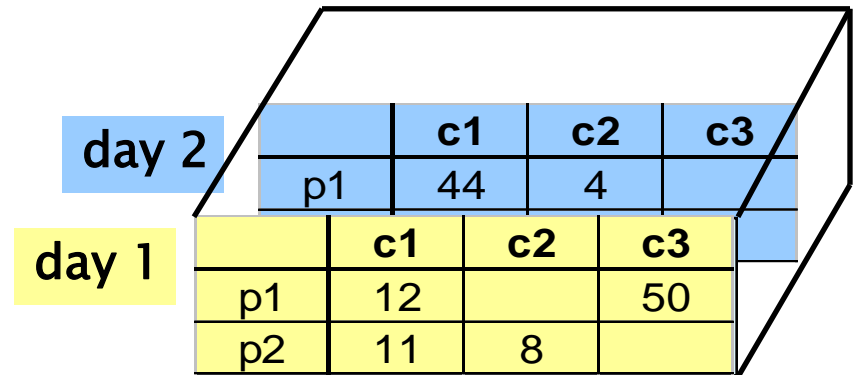


3-D Cube

Fact table view:

sale	prold	storeld	date	amt
	p1	c1	1	12
	p2	c1	1	11
	p1	c3	1	50
	p2	c2	1	8
	p1	c1	2	44
	p1	c2	2	4

Multi-dimensional cube:

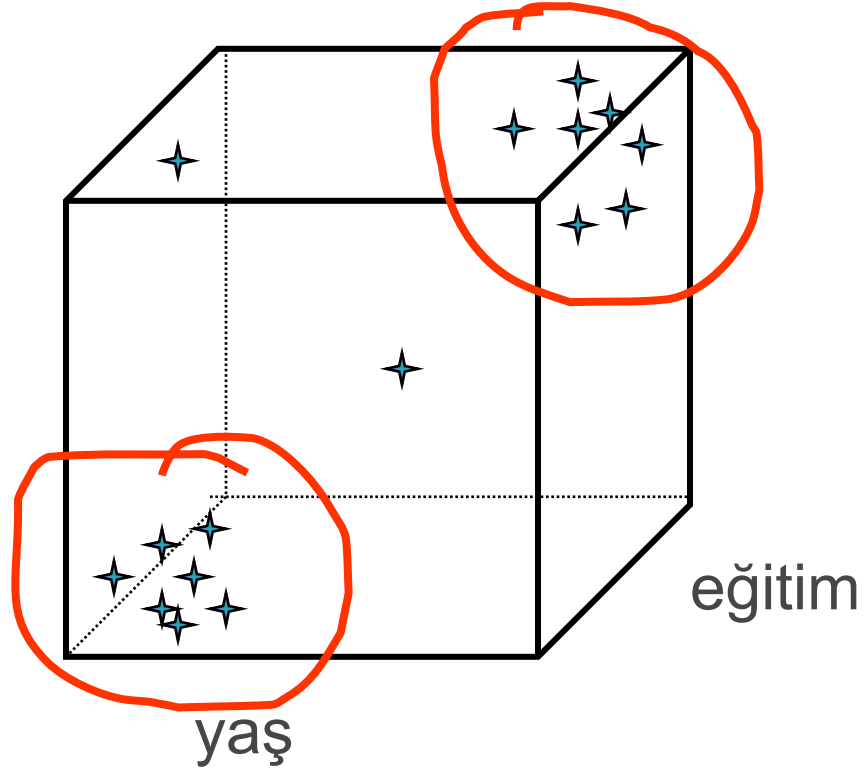


dimensions = 3

□ **Veri madenciliği:**

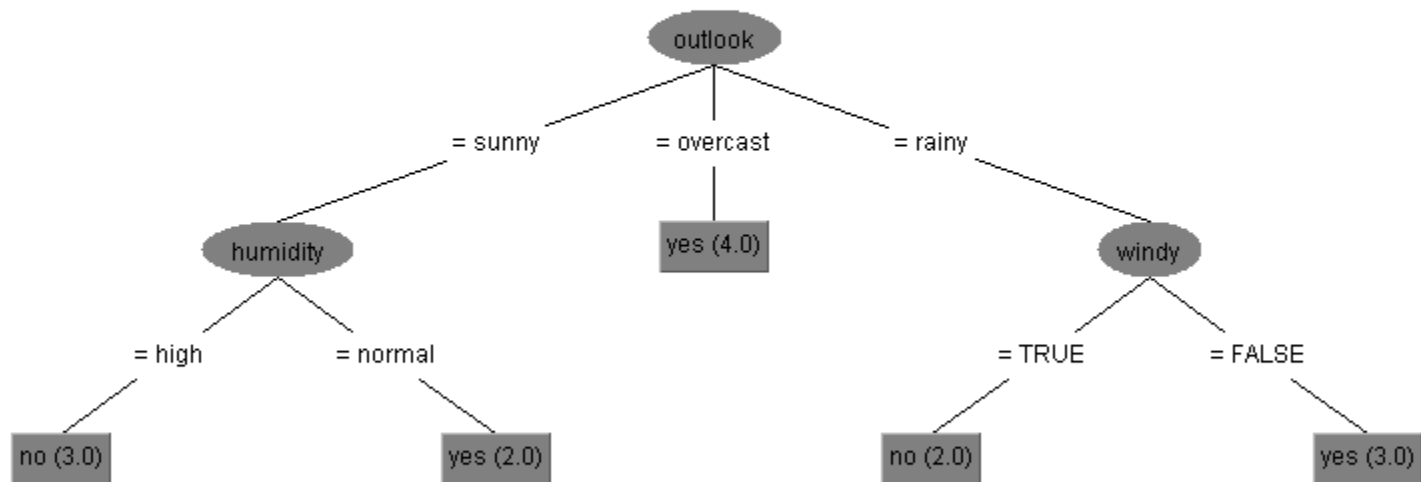
- **Sınıflandırma**
- **Gruplandırma**
- **Kümeleştirme**
- **İlişkili kurallar çıkarımı**

gelir

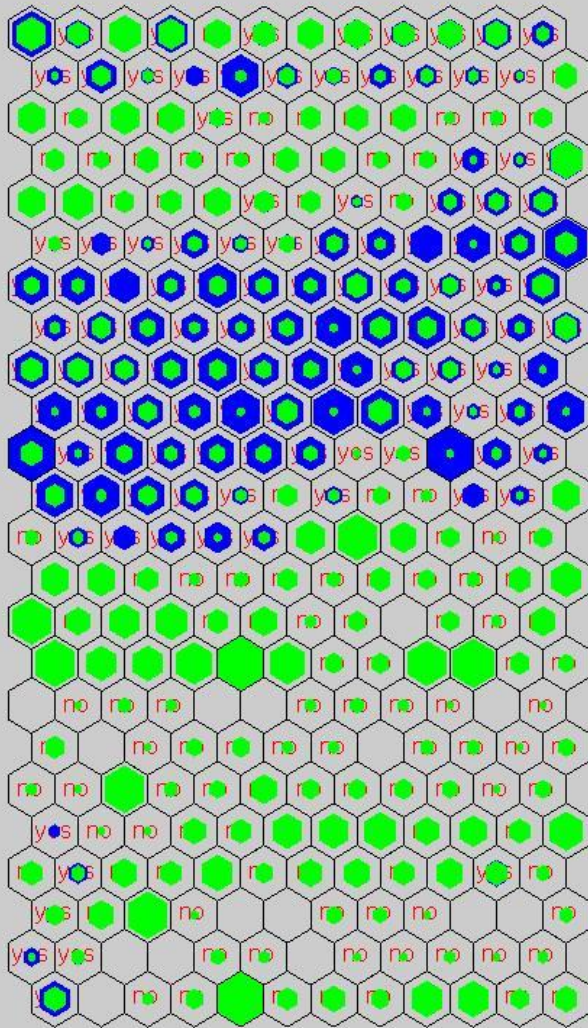




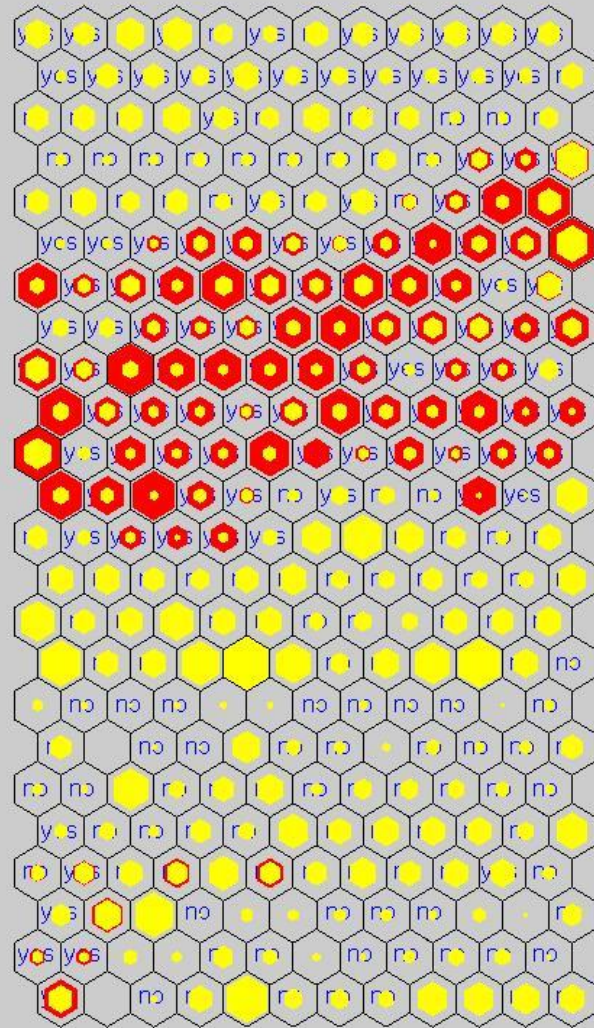
Tree View

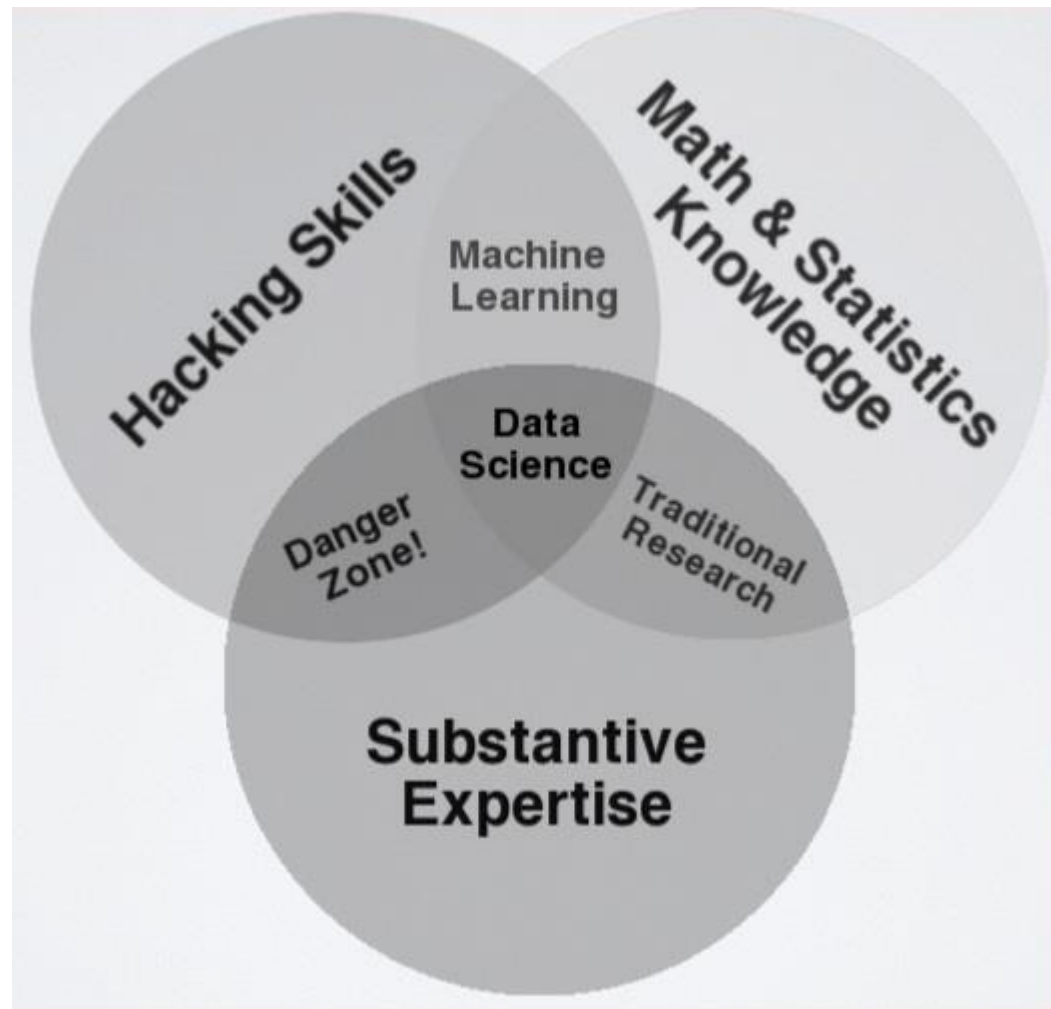


Train



Test





«You're already a walking
sensor platform»

CIA CTO - Ira Gus Hunt

- **KİM**
- **NERDE**
- **NASIL**
- **NE İÇİN**

SORULAR

ibrahim.zincir@yasar.edu.tr