**Hafta 7**

**IV. Aşama - Model Kurma:** Verinin karakterize edilmesinin ardından bu veri setini temsil edecek bir model kurmanızı istiyoruz. İnternet üzerinde hazır birçok modeli doğrudan uygulayabilirsiniz. Burada herhangi bir kısıtlama yoktur. İstediğiniz uygulamayı, istediğiniz platformu kullanabilirsiniz. Özellikle bu dönem Yapay Zeka dersinde elde ettiğiniz kazanımlar bu aşamada size önemli bir katkı sunacaktır.

Veri modeli verinin hangi kurallara göre yapılandırıldığını analiz edip bunları belirlemeye denir. Burada elde edilen verilere bakarak tahminler yürütmek analizler yapmak mümkündür.

Veri Modellerinin Sınıflandırılması:

**Veri Modellerinin Sınıflandırılması**

1. **Yüksek Seviyeli Veri Modelleri:**Bu modeller nesneler(entities), nitelikler(attributes) ve ilişkiler(relationships) gibi konseptler kullanırlar. Bu modellere örnek; EER, UML, v.b.
2. **Mantıksal Veri Modelleri:** Bu modeldeki konseptler verilerin bilgisayarda fiziksel olarak organize olma biçimine çok benzerler. Bunlara örnekler; ilişkisel, nesneye-yönelik, nesne-ilişkisel, XML veri modelleridir.
3. **Fiziksel Veri Modelleri:** Bu modeller verilerin bilgisayarda nasıl tutulduklarını detaylı olarak gösteren konseptlere sahiptirler. Bunlar kayıtların biçimiyle, kayıtların sırasıyla, ulaşma yollarıyla ilgili bilgiler içerirler. Dizinleme (B+ tree) fiziksel veri modelinin önemli bir ögesidir.

Bu verilerden model oluşturmak için geliştrilirmiş çeşitli yöntemler vardır. Bu yöntemleri incelemeden önce regrasyon kavramını irdelememiz gerekiyor.

**Regresyon**, bir bağımlı değişken (genellikle Y ile ifade edilir) ile bir dizi diğer değişken (bağımsız değişkenler -X- olarak bilinir) arasındaki ilişkinin gücünü belirlemeye çalışan finans, yatırım ve diğer disiplinlerde kullanılan istatistiksel bir ölçümdür.Regresyon, yatırım ve finansal danışmanların varlıklara değer vermelerine ve [**emtia**](https://finanswebde.com/emtia) fiyatları ve bu emtialarda işlem yapan işletmelerin hisse senetleri gibi değişkenler arasındaki ilişkileri anlamalarına yardımcı olur. Daha karmaşık veri ve analiz için doğrusal olmayan regresyon yöntemleri olmasına rağmen iki temel regresyon türü öne çıkar: doğrusal regresyon ve çoklu doğrusal regresyon.

**Veri Modelleme Teknikleri**

**1 – Varlık İlişkisi Şeması**  
  
•Varlık-ilişki modeli, ya da kısaca E-R modeli (Entity-Relationship model) 1976 yılında P.P. Chen tarafından geliştirilen bir modeldir.

•Bugüne kadar varlık-ilişki modeline dayalı hiçbir VTYS geliştirilmemiştir.

•Buna karşılık varlık-ilişki modeli, VTYS’den bağımsız veri çözümlemede ve semantik veri modellemede en çok kullanılan modeldir.

•Bu model kullanılarak önce;

–VTYS’den bağımsız olarak veriler çözümlenir,

–Veri modellemesi yapılır,

–Veriler ve veriler arası ilişkilerin anlamları ve özellikleri incelenerek E-R çizelgeleri oluşturulur;

–Kullanılacak VTYS belirlenir sonra da E-R çizelgeleri bu sistemin veri modeline dönüştürülerek veri tabanı şemaları oluşturulur.

**2 – Data Matrisi**  
Bir Data Matrisi data modeli hakkında daha detaylı bilgi sağlar ve farklı formların çeşitliliğini kapsar. Tipik olarak bir Data Matrisi iş tablosu formatında tutulur ve her bir nitelik hakkında ek bilgiyle birlikte niteliklerin bir listesini içerir. Bir data matrisi sütununda tutulabilen bazı genel ek bilgi tipleri aşağıdakileri içerir:  
• Data tipi  
• Uygun değerler  
**3 – Data Planlama Verisi**  
Bir Data Planlama Verisi, bilginin birbirine bağlı iki farklı veri tabanında nasıl stoklandığını gösterir. Veri tabanları genellikle organizasyonlar, organizasyon ve üçüncü parti tedarikçi veya ikili işbirliği organizasyonlarının sahip olduğu iki farklı bilgi teknolojisi sisteminin parçasıdır.  
Data değişimi ve ithali için aynı anda iki sisteme birden bağlanılacaksa, data planlama verisi iyi bir seçim olacaktır.  
  
**4 – Data Akış Şeması**  
Data Akış Şemaları BABOK Guide kitabında listelenmiş iki formel seçenekten birini veya basit bir iş akışı şemasını kullanarak oluşturulabilir.  
Data Akış Şeması hangi datanın oluşturulduğunu veya sistem tarafından tasarlandığını veya tutulduğunu size söylemez. Bir Data Akış Şeması data stoklarını, süreçlerini ve data çıktılarını gösterir.  
Eğer data çok fazla süreçten geçiyorsa, ne zaman ve nasıl bu süreçlerin gerçekleştirileceğini belirlemeye yardımcı olduğu için Data Akış Şeması iyi bir seçimdir. O halde her data stoku Data Planlama Verisi’ni kullanan her süreci ve ERD ve/veya Data Matrisini kullanarak modellenebilir.