**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.

**Regrasyon:** Regresyon, bağımsız öngörücülere dayanan bir hedef değeri modelleme yöntemidir. Bağımlı değişken olarak da bilinen sonuç değişkeni ile genellikle bağımsız değişken olarak adlandırılan bir veya daha fazla değişken arasındaki ilişkiyi bulmak için kullanılan istatistiksel bir araçtır.  Bu yöntem çoğunlukla değişkenler arasındaki neden sonuç ilişkisini tahmin etmek ve bulmak için kullanılır. **Regresyon**teknikleri çoğunlukla bağımsız değişkenlerin sayısına ve bağımsız ve bağımlı değişkenler arasındaki ilişkinin türüne göre farklılık gösterir.

**Regresyon Analizi Nedir?**

**Regresyon analizi** bağımlı bir değişken ile bağımsız bir değişken arasındaki ilişkinin ortadan kaldırılması için kullanılan istatistiksel bir yöntemdir. Değişkenler arasındaki ilişkinin gücüne erişmede faydalıdır.

Değişkenler arasındaki gelecekteki ilişkinin modellenmesine de yardımcı olur. Regresyon analizi doğrusal, doğrusal olmayan ve çoklu doğrusal gibi çeşitli tiplerde yapılır. Ancak en faydalı olanlar basit doğrusal ve çoklu doğrusaldır.

Bununla birlikte, doğrusal olmayan analiz temel olarak karmaşık veri kümeleriyle başa çıkmada yardımcı olur. Karmaşık veri setlerinde bağımsız ve bağımlı değişken doğrusal olmayan bir ilişki gösterir.

**Veri Modelleme Teknikleri**

**1 – Varlık İlişkisi Şeması**  
  
Bir Varlık İlişkisi Şeması (Entity Relationship Diagram – ERD), ilişkileri ve nitelikleri modeller. ERD’ler, düşük seviye fiziksel veri tabanı modelinden, orta seviye makul veri tabanı ve üst seviye iş alanı modeline kadar çoklu seviyelerindeki data modellerinde kullanılabilir. Eğer çoklu konsept ve veritabanı şeması varsa ve her konsept ya da şemanın sınırları analiz edilyorsa, ERD iyi bir seçenektir. Nitelikleri tanımlayarak her varlığın neye ait olduğunu tespit edersiniz. İlişkileri belirleyerek her varlığın modelinizdeki diğer varlıklara nasıl bağlı olduğunu tespit edersiniz.  
  
**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  
• Gerekli ve opsiyonel  
• Örnek Data  
• Notlar  
Data modelindeki her bir nitelik hakkında detaylı bilgileri analiz etmek gerektiğinde Data Matrisi iyi bir seçenektir. Bu bilgi genelde fiziksel veri tabanını dizayn etmek ve tasarlamak için kullanılır ve bu sebeple data mimarı veya veri tabanı geliştiricileri buna ihtiyaç duyar.

**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.