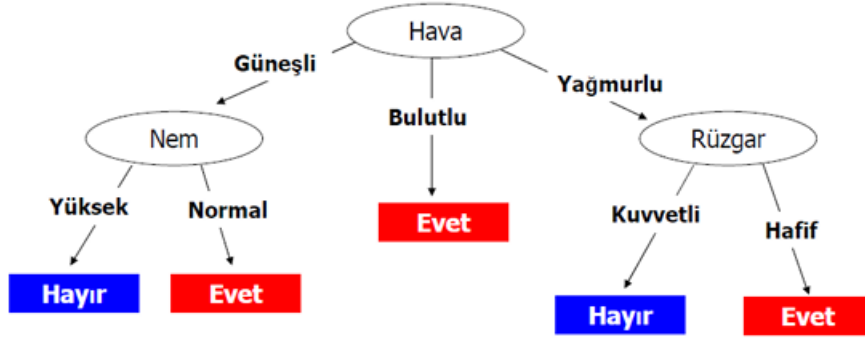


KARAR AĞAÇLARI

Karar ağacı, kararlar veya tahminler yapmak için kullanılan akış şemasına benzer bir yapıdır. Niteliklere ilişkin kararları veya testleri temsil eden düğümlerden, bu kararların sonuçlarını temsil eden dallardan ve nihai sonuçları veya tahminleri temsil eden yaprak düğümlerden oluşur. Her bir iç düğüm bir nitelik üzerindeki teste karşılık gelir, her dal testin sonucuna karşılık gelir ve her yaprak düğüm bir sınıf etiketine veya sürekli bir değere karşılık gelir.

Karar ağacı algoritması, denetimli öğrenme kategorisine girer. Hem regresyon hem de sınıflandırma problemlerini çözmek için kullanılır.

Karar ağaçları metodu, giriş verisinin bir algoritma yardımıyla gruplara bölünerek tüm elemanlarının aynı sınıf etiketine sahip olması için yapılan sınıflama işlemidir. Giriş verisinin bir kümeleme algoritması yardımıyla tekrar tekrar gruplara bölünmesine dayanır. Grubun tüm elemanları aynı sınıf etiketine sahip olana kadar kümeleme işlemi derinlemesine devam eder.



Karar Ağacı Yapısı:

Kök Düğüm: Veri setinin tamamını ve verilecek ilk kararı temsil eder.

Dahili Düğümler: Niteliklere ilişkin kararları veya testleri temsil eder. Her iç düğümün bir veya daha fazla dalı vardır.

Dallar: Başka bir düğüme giden bir kararın veya testin sonucunu temsil eder.

Yaprak Düğümleri: Nihai kararı veya tahmini temsil eder. Bu düğümlerde başka bölünme meydana gelmez.

Nasıl Çalışır?

Öznitelik Ölçüm Seçimi: **Gini safsızlığı**, **entropi** veya **bilgi kazancı** gibi bir ölçüm kullanılarak verileri bölmek için en iyi özellik seçilir.

Veri setini Bölme: Elimizdeki veri seti seçtiğimiz özniteliğe göre alt setler oluşturulur

Sürecin Tekrarlanması: Süreç, her bir alt küme için yinelemeli olarak tekrarlanarak, bir durdurma kriteri karşılanana kadar yeni bir iç düğüm veya yaprak düğüm oluşturulur.

1) Öğrenme kümesi oluşturulur.

2) Öğrenme kümesindeki örnekleri en iyi ayıran nitelikler belirlenir.

3) Seçilen nitelikler ile ağacın düğümleri oluşturulur ve her bir düğümde alt düğümler veya ağacın yapraklarını oluşturulur. Alt düğümlere ait alt veri kümesinin örneklerini belirlenir.

4) 3. adımda oluşturulan her alt veri kümesi için:

- Örneklerin hepsi aynı sınıfa aitse
- Örnekleri bölecek nitelik kalmamışsa
- Kalan niteliklerin değerini taşıyan örnek yoksa işlemi sonlandır. Diğer durumda alt veri kümesini ayırmak için 2. adımdan devam edilir.

Ezber (Overfitting):

Tüm makine öğrenmesi yöntemlerinde verinin ana hatlarının modellenmesi esas alındığı için öğrenme modelinde ezberden (overfitting) kaçınılmalıdır. Tüm karar ağaçları önlem alınmazsa ezber yapar. Bu yüzden ağaç oluşturulurken veya oluşturulduktan sonra budama yapılmalıdır.

Budama (Pruning)

Aşırı uyumun üstesinden gelmek için budama teknikleri kullanılır. Budama, örnekleri sınıflandırmada çok az güç sağlayan düğümleri kaldırarak ağacın boyutunu küçültür. Budamanın iki ana türü vardır:

1.Ön budama (Erken Durdurma): Belirli kriterleri (örneğin maksimum derinlik, yaprak başına minimum örnek sayısı) karşıladığında ağacın büyümesini durdurur.

2.Budama sonrası: Tamamen büyümüş bir ağaçtan önemli bir güç sağlamayan dalları kaldırır.

