Karar Ağaçları algoritmasını bir örnek üzerinde uygulamaya çalışalım. Algoritmayı hatırlamak gerekirse ;

1)T öğrenme kümesini oluştur.

2)T kümesindeki örnekleri en iyi ayıran niteliği belirle.

3)Seçilen nitelik ile ağacın bir düğümünü oluştur ve bu düğümden alt düğümleri veya yaprak düğümünü oluştur. Alt düğümlere ait alt veri kümesinin örneklerini incele.

4)Adımda oluşturulan her alt veri kümesi için

-Örneklerin hepsi aynı sınıfa ait

-Örnekleri bölecek özellik kalmamış

-Kalan özelliklerin değerini taşıyan örnek kalmayıncaya kadar devam eder. 2. adımdaki örnekleri en iyi ayıran nitelik Bilgi Kazanımı ile seçilir.

Değişken Seçimi Ölçüsü: Bilgi Kazanımı

- En yüksek bilgi kazanımını veren değişkeni seç

-S, Ci sınıfından si satır içerir. i = {1, …, m}

-Herhangi bir satiri sınıflandırmak için gereken **bilgi**



Bir A değişkeninin {a1,a2,…,av} değerleri ile düzensizliği (**entropi**)



Bir A değişkeni kullanılarak ağacın dallanmasıyla kazanılan bilgi



Basit bir data üzerinden örnek verecek olursak;

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Yaş | Gelir | Öğrenci | Kredi Durumu | Bilgisayar Durumu |
| <30 | Yüksek | Hayır | Vasat | Hayır |
| <30 | Yüksek | Hayır | Mükemmel | Hayır |
| 31----40 | Yüksek | Hayır | Vasat | Evet |
| >40 | Orta | Hayır | Vasat | Evet |
| >40 | Düşük | Evet | Vasat | Evet |
| >40 | Düşük | Evet | Mükemmel | Hayır |
| 31----40 | Düşük | Evet | Mükemmel | Evet |
| <30 | Orta | Hayır | Vasat | Hayır |
| <30 | Düşük | Evet | Vasat | Evet |
| >40 | Orta | Evet | Vasat | Evet |
| <30 | Orta | Evet | Mükemmel | Evet |
| 31---40 | Orta | Hayır | Mükemmel | Evet |
| 31---40 | Yüksek | Evet | Vasat | Evet |
| >40 | Orta | Hayır | Mükemmel | Hayır |

P sınıfı: Bilgisayar durumu= “evet”

N sınıfı: Bilgisayar durumu= ”hayır”

I(p, n) = I(9, 5) =0.940

Yaş için entropiyi hesaplayalım:

E(yaş)= 5/14 I(2,3)+4/14 I(4,0)+5/14 I(3,2)=0.694

Kazanç(yaş)= I(p,n)-E(yaş)=0,246

Buna benzer

Kazanç(gelir)=0.029

Kazanç(öğrenci)=0.151

Kazanç(Kredi durumu)=0.048

Bu durumda en yüksek kazanca sahip olan yaş kök düğüm olur.

Yaş

<30 31…40 >40