

KARAR DESTEK SİSTEMLERİ

KARAR AĞAÇLARI VE UYGULAMASI

Oğuzhan OĞUZ
Yönetim Bilişim Sistemleri 4.Sınıf N.Ö
10010207007

Karar Ağaçları

- Verilerin analizinde istatistik, pek çok problemi çözmede önemli bir araçtır. Ancak, bazı durumlarda kullanımı sınırlıdır. Bu durumlarda, yapay sinir ağları, kural çıkarımı, mantık programlama, karar ağaçları, genetik algoritma gibi akıllı veri analiz yöntemlerini içeren veri madenciliğine gereksinim ortaya çıkmaktadır . Ancak, bu yöntemlerden bazıları **kara kutu** yaklaşımları şeklindedir.
- Başka bir ifade ile açıklama getiremeden iyi tahmin yaparlar. Bu da, bu yöntemlerin zayıf yönünü oluşturur.
- Karar ağacı yaklaşımı ise bu zayıflığın söz konusu olmadığı veri madenciliği yöntemlerinden biridir

Karar Ağaçları

- Karar ağacı yaklaşımı, hedef fonksiyonlarını yaklaşık olarak hesaplamak için kullanılan ve öğrenme fonksiyonunun karar ağacı ile gösterildiği bir yöntemdir.
- Bir karar ağacı ise ağaç görünümünde tanımlayıcı ve tahmin edici bir modeldir .
- Bu model, karar alıcıya karar alırken hangi faktörlerin göz önüne alınması ve her bir faktörün kararın farklı çıktıları ile geçmişte nasıl ilişkili olduğunun belirlenmesi konularında yardımcı olur.

Karar Ağacı Algoritması

Karar Ağacı Algoritması 2 aşamadan oluşmaktadır;

1. Ağacı Oluşturma

- En başta bütün öğrenme kümesi ağaçtır.

2 . Ağacı Budama

- Öğrenme kümesindeki gürültülü verilerden oluşan ve test kümesinde hataya neden olan dallar silinir.

ÖNEMLİ NOT

- Karar Ağacı iyi bir çözümdür ancak optimum değildir.
- Optimum bir karar ağacının oluşturulması için bir kuralın olması gerekir.

Karar Ağacı Algoritması Adımları

ADIM 1: Karar ağacının hangi kararı alacağı belirlenir.

- Örnek veri setinde o hafta hava, ebeveyn ve para duruma göre ne yapılacağına karar verilecektir.

Karar Ağacı Algoritması Adımları



ADIM 2 : Sistemin ENTROPY'si hesaplanır.

Entropy formülü şu şekilde dir;

n

$$E(S_1, S_2, \dots, S_m) = - \sum_{i=1}^m \frac{S_i}{S} \log_2 \left(\frac{S_i}{S} \right)$$

m=sınıf sayısı

m=sınıf sayısı

Karar Ağacı Algoritması Adımları

ADIM 3 : Ağacın en üstünde yani kökte(root) konumlanacak özellik belirlenir.

Peki neye göre belirlenir?

- Bilgi kazancı (information gain) en yüksek olan özellik ağacın en üstünde konumlandırılır.

Entropy ve Bilgi Kazancı

- **Bilgi Kazancı (Information Gain):** ID3, C4.5 gibi karar ağacı metotlarında en ayırt edici niteliği belirlemek için her nitelik için **bilgi kazancı** ölçülür.
- Bilgi Kazancı ölçümünde **Entropy** kullanılır.
- **Entropy** rastgeleliğin, belirsizliğin ve beklenmeyen durumun ortaya çıkma olasılığını gösterir.

Bilgi Kazancı Nasıl Hesaplanır?

▣ A özelliğinin , S örneği için bilgi kazancı :

$$\text{Kazanç}(S, A) = \text{Entropy}(S) - \sum_{v \in A} \text{Entropy}(P_v)$$

$v = A$ 'nın değerleri $v = A$ 'nın değerleri

Karar Ağacı Algoritması Adımları

- Her özellik için ayrı ayrı bilgi kazancı hesaplanır ve bilgi kazancı en yüksek olanlar kök olarak alınır.
- Bu işlemler her düğüm için aşağıdaki durumlardan biri oluşuncaya kadar devam eder.
 - ✓ Örneklerin hepsi aynı sınıfa ait
 - ✓ Örnekleri bölecek özellik kalmamış
 - ✓ Kalan özelliklerin değerini taşıyan örnek yok

Karar Ağacı Algoritması Adımları

- Ve bu adımlar tamamlanarak karar ağacı oluşturulmuş olur.

UYGULAMA

- Öncelikle örnek veri setimiz belirlenen haftada hava, ebeveyn ve para duruma göre ne yapılacağına karar verecektir.
- Örnek veri setimizi inceleyelim;

UYGULAMA

	A	B	C	D	E	F	
1							
2							
3		HAFTALAR	Hava	Ebeveyn	Para	KARAR	
4		1.Hafta	Güneşli	Var	Var	Sinema	
5		2.Hafta	Güneşli	Yok	Var	Tenis	
6		3.Hafta	Rüzgarlı	Var	Var	Sinema	
7		4.Hafta	Yağmurlu	Var	Yok	Sinema	
8		5.Hafta	Yağmurlu	Yok	Var	Evde kal	
9		6.Hafta	Yağmurlu	Var	Yok	Sinema	
10		7.Hafta	Rüzgarlı	Yok	Yok	Sinema	
11		8.Hafta	Rüzgarlı	Yok	Var	Alışveriş	
12		9.Hafta	Rüzgarlı	Var	Var	Sinema	
13		10.Hafta	Güneşli	Yok	Var	Tenis	
14							
15							

UYGULAMA

- Veri setimizde toplam 10 örnek vardır. Bu 10 örnekten;
 - ✓ 6 örnek için karar sinema($6/10$)
 - ✓ 2 örnek için karar tenis oynamak($2/10$)
 - ✓ 1 örnek için karar evde kalmak($1/10$) ve
 - ✓ 1 örnek için karar alışverişe gitmek($1/10$) olduğuna göre Entropy değerini hesaplayalım.

UYGULAMA

- Entropy değeri =

$$- (6/10) \log_2(6/10) - (2/10) \log_2(2/10) - (1/10) \log_2(1/10) - (1/10) \log_2(1/10)$$

$$\underline{E(S)=1,571}$$

UYGULAMA

- Sistemimizin Entropy değerini bulduk. Şimdi kök özelliğimizin hangisinin olması gerektiğine karar vermek için özelliklerimiz olan ; hava, ebeveyn ve para özelliklerimizin bilgi kazançlarını hesaplayıp en yüksek kazanca sahip olan özelliği kök özellik olarak konumlandırmamız gerek.

UYGULAMA

- Hava özelliğimiz için Bilgi Kazancı değerimiz;
 $\text{Bilgi Kazancı}(S, \text{Hava Durumu}) = ?$
- Şimdi hava durumunun bilgi kazancı için hava durumu özelliğinin de Entropy değerine ihtiyacımız var o yüzden hava durumunun önce Entropy değerini hesaplayalım;
- Hava durumunun alt özellikleri: güneşli, rüzgarlı ve yağmurlu idi.

UYGULAMA

- ▣ Güneşli=3 (bunlardan 1'i sinema, 2'si tenis)
- : Rüzgarlı=4 (bunlardan 3'ü sinema, 1'i alışveriş)
- : Yağmurlu=3 (bunlardan 2'si sinema, 1'i evde kalmak)
- : Öyleyse Entropy değerleri;
- ✓ $E(S_{\text{güneşli}}) = - (1/3) \log_2 (1/3) - (2/3) \log_2 (2/3) = 0,918$
- ✓ $E(S_{\text{rüzgarlı}}) = - (3/4) \log_2 (3/4) - (1/4) \log_2 (1/4) = 0,811$
- ✓ $E(S_{\text{yağmurlu}}) = - (2/3) \log_2 (2/3) - (1/3) \log_2 (1/3) = 0,918$

UYGULAMA

- Bu Entropy değerlerini Bilgi Kazancı formülünde yerine koyarsak hava durumunun bilgi kazancını buluruz.
- $\text{Bilgi Kazancı}(S, \text{Hava Durumu}) =$
 $\text{Entropy}(S) - ((\text{Bilgi Kazancı}(\text{güneşli}) \text{Entropy}(S_{\text{güneşli}})$
 $+ \text{Bilgi Kazancı}(\text{rüzgarlı}) \text{Entropy}(S_{\text{rüzgarlı}}) + \text{Bilgi}$
 $\text{Kazancı}(\text{yağmurlu}) \text{Entropy}(S_{\text{yağmurlu}}))$
- $= 1,571 -$
 $((3/10) * 0,918 + (4/10) * 0,811 + (3/10) * 0,918)$
- $\text{Bilgi Kazancı}(S, \text{Hava Durumu}) = 0,70$

UYGULAMA

- Hava durumunun bilgi kazancını bulduktan sonra sırada ebeveyn durumunun bilgi kazancını bulmak var. Onun özellikleri de ; var ve yok. Bu 2 özelliğin öncelikle entropy değerlerini bulup oradan bilgi kazancı formülünde yerine koyacağız.
- Bilgi Kazancı(S,Ebeveyn Durumu)=?

UYGULAMA

☐ Ebeveyn durumunun alt özellikleri;

: Var=5 (5'i de sinema)

: Yok=5 (2'si tennis, 1'i sinema, 1'i alışveriş, 1'i de evde kalmak)

: Öyleyse Entropy değerleri;

$$✓ E(s_{var}) = -(5/5) \log_2(5/5) - (5/5) \log_2(5/5) = 0$$

$$✓ E(s_{yok}) = -(2/5) \log_2(2/5) - (2/5) \log_2(2/5) - (1/5) \log_2(1/5) - (1/5) \log_2(1/5) = 2,922$$

UYGULAMA

- Bu Entropy değerlerini Bilgi Kazancı formülünde yerine koyarsak ebeveyn durumunun bilgi kazancını buluruz.
- Bilgi Kazancı(S,Hava Durumu)=
- $\text{Entropy}(S) - ((\text{Bilgi Kazancı}(\text{var})\text{Entropy}(S_{\text{var}}) + \text{Bilgi Kazancı}(\text{yok})\text{Entropy}(S_{\text{yok}}))$
- $= 1,571 - ((5/10) * 0 + (5/10) * 1,922)$
- Bilgi Kazancı(S,Ebeveyn Durumu)=0,61

UYGULAMA

- Ebeveyn durumunun bilgi kazancını bulduktan sonra sırada para durumunun bilgi kazancını bulmak var. Onun özellikleri de ; var ve yok. Bu 2 özelliğin öncelikle entropy değerlerini bulup oradan bilgi kazancı formülünde yerine koyacağız.
- $\text{Bilgi Kazancı}(S, \text{Para Durumu}) = ?$

UYGULAMA

- ▣ Para durumunun alt özellikleri;
- : $Var=7$ (3'ü sinema, 2'si tenis, 1'i alışveriş, 1'i evde kalmak)
- : $Yok=3$ (3'ü de sinema)
- : Öyleyse Entropy değerleri;
- : $E(s_{var})=1,842$
- : $E(s_{yok})= 0$

UYGULAMA

- Bu Entropy değerlerini Bilgi Kazancı formülünde yerine koyarsak para durumunun bilgi kazancını buluruz.
- Bilgi Kazancı(S,Para Durumu)=
- $\text{Entropy}(S) - ((\text{Bilgi Kazancı}(\text{var})\text{Entropy}(S_{\text{var}}) + \text{Bilgi Kazancı}(\text{yok})\text{Entropy}(S_{\text{yok}}))$
- $= 1,571 - ((5/10) * 1,842 + (5/10) * 0$
- Bilgi Kazancı(S,Para Durumu)=0,2816

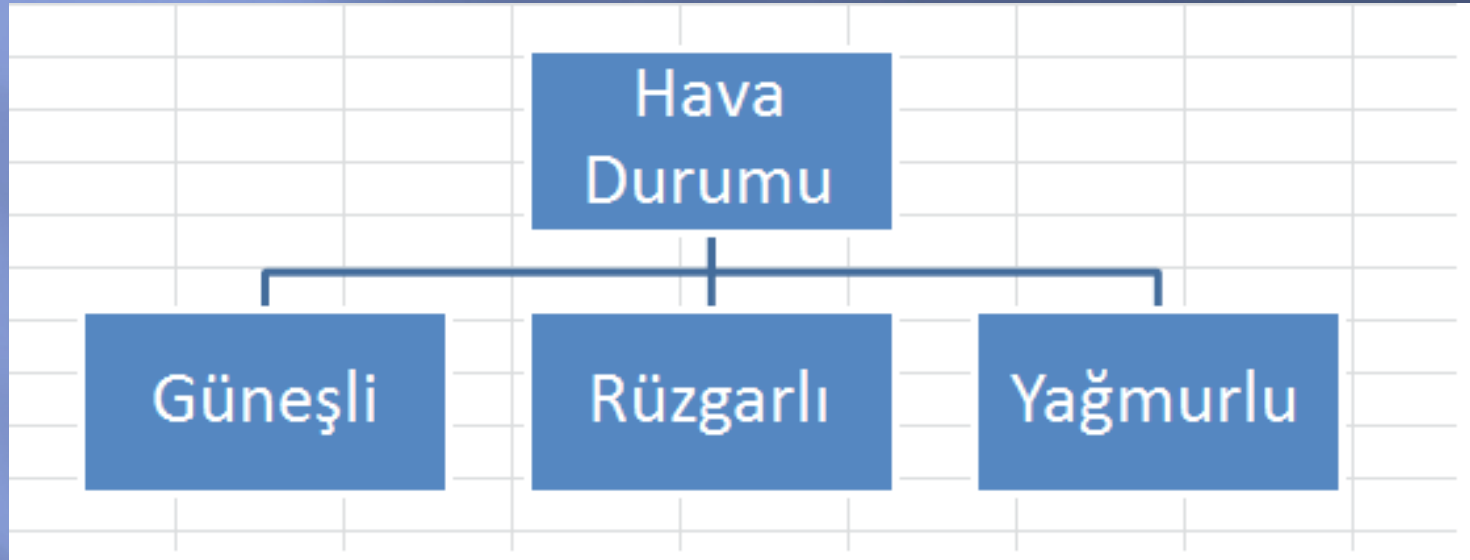
UYGULAMA

- Şimdi sıra geldi 3 özelliğin bilgi kazançlarını karşılaştırıp kazancı en yüksek olan özelliği kök özellik olarak konumlandırmaya.
- **Bilgi Kazancı(S,Hava Durumu)=0,70**
- Bilgi Kazancı(S,Ebeveyn Durumu)=0,61
- Bilgi Kazancı(S,Para Durumu)=0,2816
- Görüldüğü gibi bilgi kazancı en yüksek olan HAVA DURUMU bizim kök özelliğimiz olmuştur.

UYGULAMA

- Hava durumu bizim ilk kök hücremiz oldu fakat daha 2 özelliğimiz var ebeveyn durumu ve para durumu. Bunlardan hangisinin kök özellik olacağına tekrar bilgi kazançları hesaplanarak bakılması gerekir ve bunun için hava durumunun özellikleri olan güneşli, rüzgarlı ve yağmurlu özelliklerinin hepsi için ayrı ayrı değerlendirerek karar ağaçlarının dalları oluşturulur.

UYGULAMA



- Şu anda karar ağacımız bu aşamadadır. Şimdi onu dallandırmaya devam ediyoruz.

UYGULAMA

- Bunun için ise hava durumunun özelliklerini tek tek ele almamız gerekli. Öncelikle güneşli özelliğinden başlayalım.

HAFTALAR	Hava	Ebeveyn	Para	KARAR
1.Hafta	Güneşli	Var	Var	Sinema
2.Hafta	Güneşli	Yok	Var	Tenis
10.Hafta	Güneşli	Yok	Var	Tenis

- Veri setimizden yalnızca hava durumu=güneşli olan verilerimizi aldık.

UYGULAMA

- - : Şimdi bu verileri ebeveyn ve para durumlarına göre değerlendirip hangi özelliğin kök özellik olacağına bakacağız.
 - : Güneşli havanın Entropy'sini daha önceden bulmuştuk;
 - : $E(X_{güneşli}) = -\left(\frac{1}{3}\right) \log_2\left(\frac{1}{3}\right) - \left(\frac{2}{3}\right) \log_2\left(\frac{2}{3}\right) = 0,918$
 - : O zaman bilgi kazancında yerine koyarsak sonuca ulaşabiliriz.

UYGULAMA

- Havanın güneşli olmasının özellikleri; ebeveyn durumu ve para durumudur.
- Ebeveyn durumu da 2 özellikten oluşur; var ya da yok.
- Var=1 (1'i de sinema)
- Yok=2 (2'si de tenis)
- $\text{Bilgi Kazancı}(S_{\text{güneşli}}, \text{Ebeveyn Durumu}) =$
- $0,198 - (1/3) * 0 - (2/3) * 0 = 0,198$
- $\text{Bilgi Kazancı}(S_{\text{güneşli}}, \text{Para Durumu}) =$
- $0,198 - (3/3) * 0,198 - (0/3) * 0 = 0$

UYGULAMA

- Görüldüğü gibi Ebeveyn Durumunun bilgi kazancı > Para Durumunun bilgi kazancı olduğundan güneşli havanın alt özelliği ebeveyn durumu olmuştur.
- Böylece güneşli havanın karar sistemi ortaya çıkmış oldu. Hava güneşli ve yanında ebeveyni varsa sinemaya, hava güneşli ve yanında ebeveyni yoksa tenise gidilecektir.

UYGULAMA



- Şu anda karar ağacımız bu durumda.

UYGULAMA

- Aynı uygulamaları rüzgarlı ve yağmurlu havada da yaptığımızda karar ağacımızı oluşturmuş olacağız.
- Rüzgarlı hava için veri setimizden yalnızca hava durumu=rüzgarlı olan verilerimizi aldık.

HAFTALAR	Hava	Ebeveyn	Para	KARAR
3.Hafta	Rüzgarlı	Var	Var	Sinema
7.Hafta	Rüzgarlı	Yok	Yok	Sinema
8.Hafta	Rüzgarlı	Yok	Var	Alışveriş
9.Hafta	Rüzgarlı	Var	Var	Sinema

UYGULAMA

-
- : Burada hava rüzgarlı olduğunda ebeveyn durumunun mu yoksa para durumunun mu daha fazla bilgi kazancı getireceğine bakacağız.
- : Rüzgarlı havanın Entropy'sini daha önceden bulmuştuk;
- : $E(S_{\text{rüzgarlı}}) = -\left(\frac{3}{4}\right) \log_2 \left(\frac{3}{4}\right) - \left(\frac{1}{4}\right) \log_2 \left(\frac{1}{4}\right) = 0,811$
- : O zaman bilgi kazancında yerine koyarsak sonuca ulaşabiliriz.

UYGULAMA

- Havanın rüzgarlı olmasının özellikleri; ebeveyn durumu ve para durumudur.
- Ebeveyn durumu da 2 özellikten oluşur; var ya da yok.
- Var=2(2'si de sinema)
- Yok=2(1'i sinema, 1'i alışveriş)
- Bilgi Kazancı($S_{\text{rüzgarlı}}$, Ebeveyn Durumu)= $0,811 - (2/4)*0 + (2/4)*1,922 = \underline{1,772}$
- Bilgi Kazancı($S_{\text{rüzgarlı}}$, Para Durumu)= $0,811 - (3/4)*1,842 + (1/4)*0 = \underline{0,5705}$

UYGULAMA

- İşlemleri tüm özellikleri konumlandırana dek devam ettiriyoruz.
- En sonunda karar ağacımızın yapısı karşımıza çıkıyor.

UYGULAMA

