

# Veri Madenciliği

Güz 2023

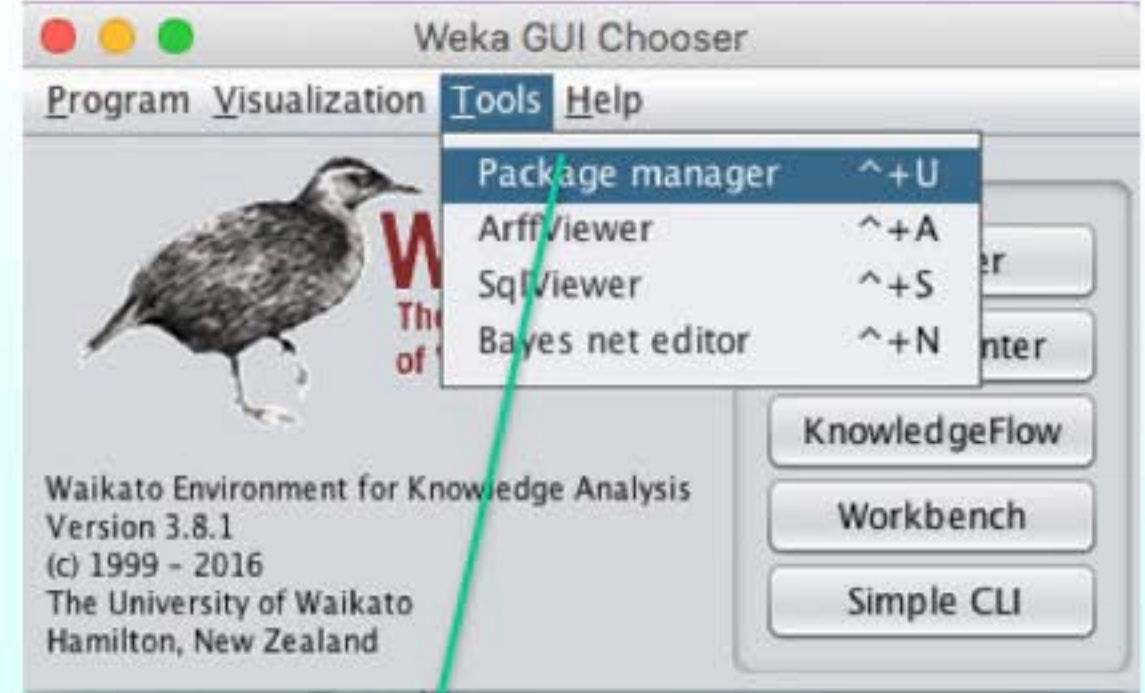
Ders 9

Çıktı: Kurallar

# İlk olarak, ihtiyacımız olan paketi yükleyin....

## EL ile Karar Ağacı Oluşturma

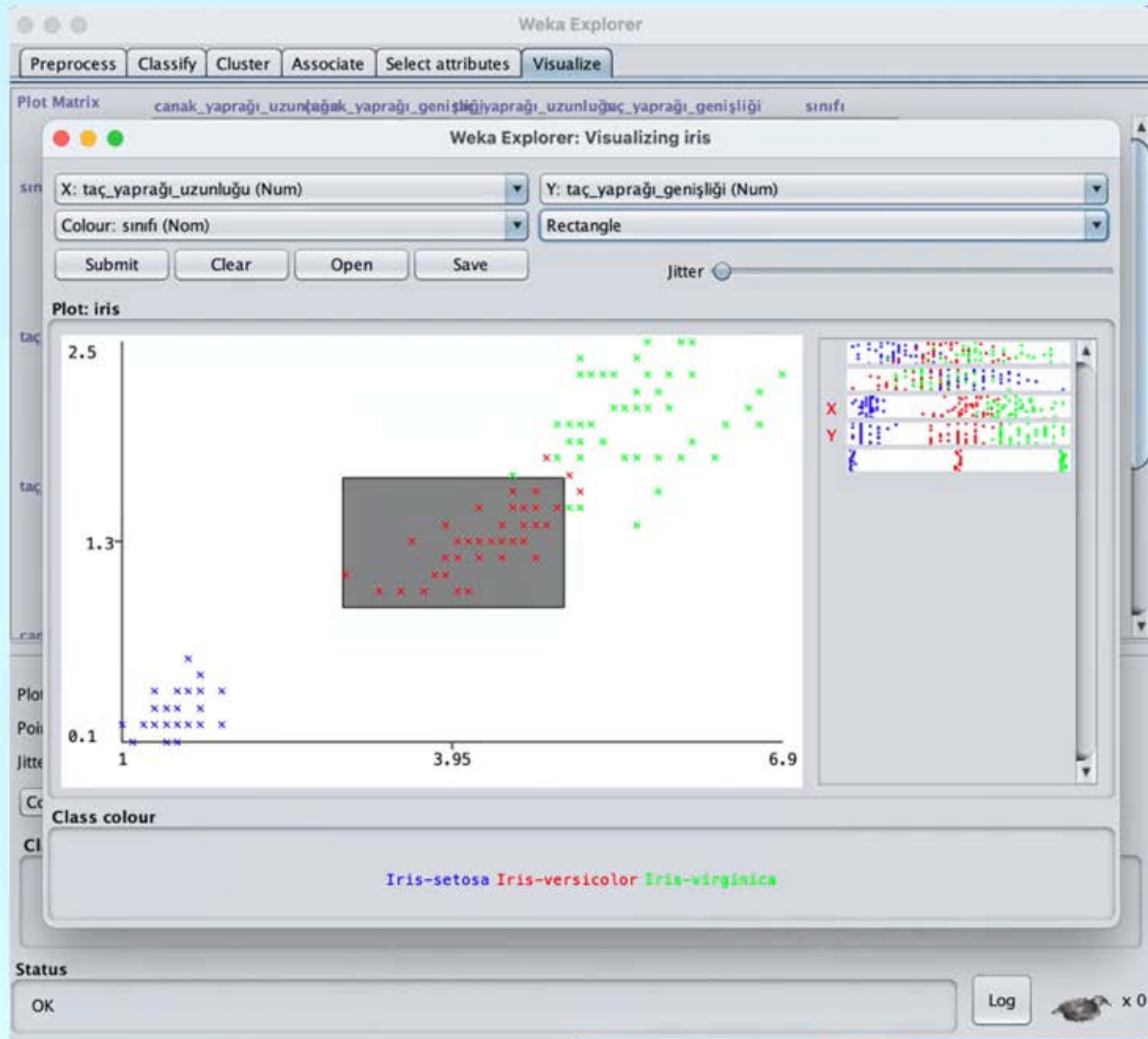
- WEKA'daki "Kullanıcı Sınıflandırma" paketini kullanmak istiyoruz. WEKA 3.8.5'de bu yüklenmelidir.
- Açılan pencereden aşağı kaydırın ve yüklemek için "userClassifier"ı seçin. Başarıyla yüklendiğini belirten bir ileti alın.





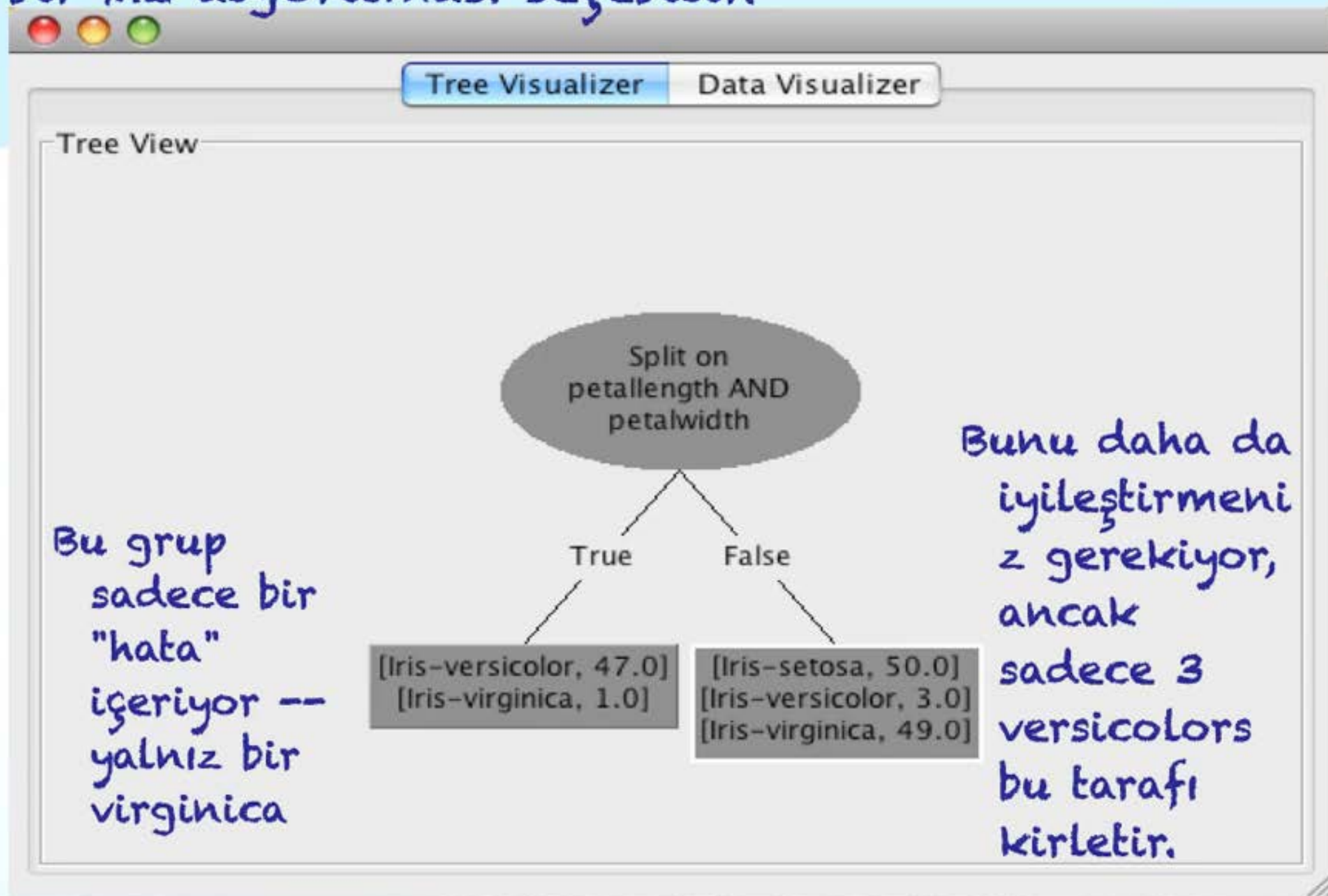
# EL ile Karar Ağacı Oluşturma

- WEKA'daki "Kullanıcı Sınıflandırma" olanağını kullanın.
- Sınıflandır sekmesine tıklayın
  - Seç/ağaçlar/KullanıcıSınıflandırması'na tıklayın
- Verileri görselleştirin.
- İki Yönlü Bölme:
  - Sınıfı iyi ayırt eden bir öznelilik gifti bulma.
  - Etraflarına çokgen çizme.
  - Sonraki slayt: Iris versicolor sınıfını "izole etmek" için taç yaprağı uzunluğu ve yaprakların genişliğini kullanın.
  - Ağacı görüntülemek için geçiş yapınız.





Bir süre el ile oluşturabilir ve sonra bitirmek için bir ML algoritması seçebilir.





# İnteraktif karar ağacı yapımı (Bunu kendi başına deneyebilirsiniz)

- Segment-challenge.arff'ı yükleyin; veri kümesini inceleyin
- *KullanıcıSınıflandırıcısı* **Seç** (ağaç sınıflandırıcısı)
- **Test kümesi** segment-test.arff kullanın
- *Veri görselleştiricisi* **ve** *ağaç görselleştiricisi* inceleyin
- *bölge-merkezli-satır* **vs** *yoğunluk-ortalaması*'nı çizin
- Dikdörtgen, çokgen ve çok çizgili seçim araçları
- ... birçok seçim ...

Ağaç görselleştiricisini sağ tıklayın ve Ağacı kabul edin

Yeterli zaman verildiğinde, veri kümesi için  
"mükemmel" bir ağaç üretebiliriz, ancak bu test  
verilerinde iyi performans gösterir mi?



# Sınıflandırma Kuralları

Öznitelik BİR ve öznitelik İKİ varsa, SINIF x'tir  
Karar ağaçlarına popüler alternatif

- **Öncül (ön koşul):** bir dizi test (tıpkı bir karar ağacının düğümlerindeki testler gibi)
- Testler genellikle mantıksal olarak (VE) birlikte sona erer (ancak genel mantıksal ifadeler de olabilir)

**Sonuç (sonuç):** sınıflar, sınıflar kümesi veya kural tarafından atanan olasılık dağılımı

- Tek tek kurallar genellikle mantıksal olarak (VEYA) birlikte sona ererken
  - Farklı sonuçlar geçerliyse gelişmeler ortaya çıkar



# Ağaçlardan Kurallara

**Kolay:** ağacı kurallar kümesine dönüştürme

- Her yaprak için bir kural:
  - Öncül, kökten yaprağa giden yoldaki her düğüm için bir koşul içerir

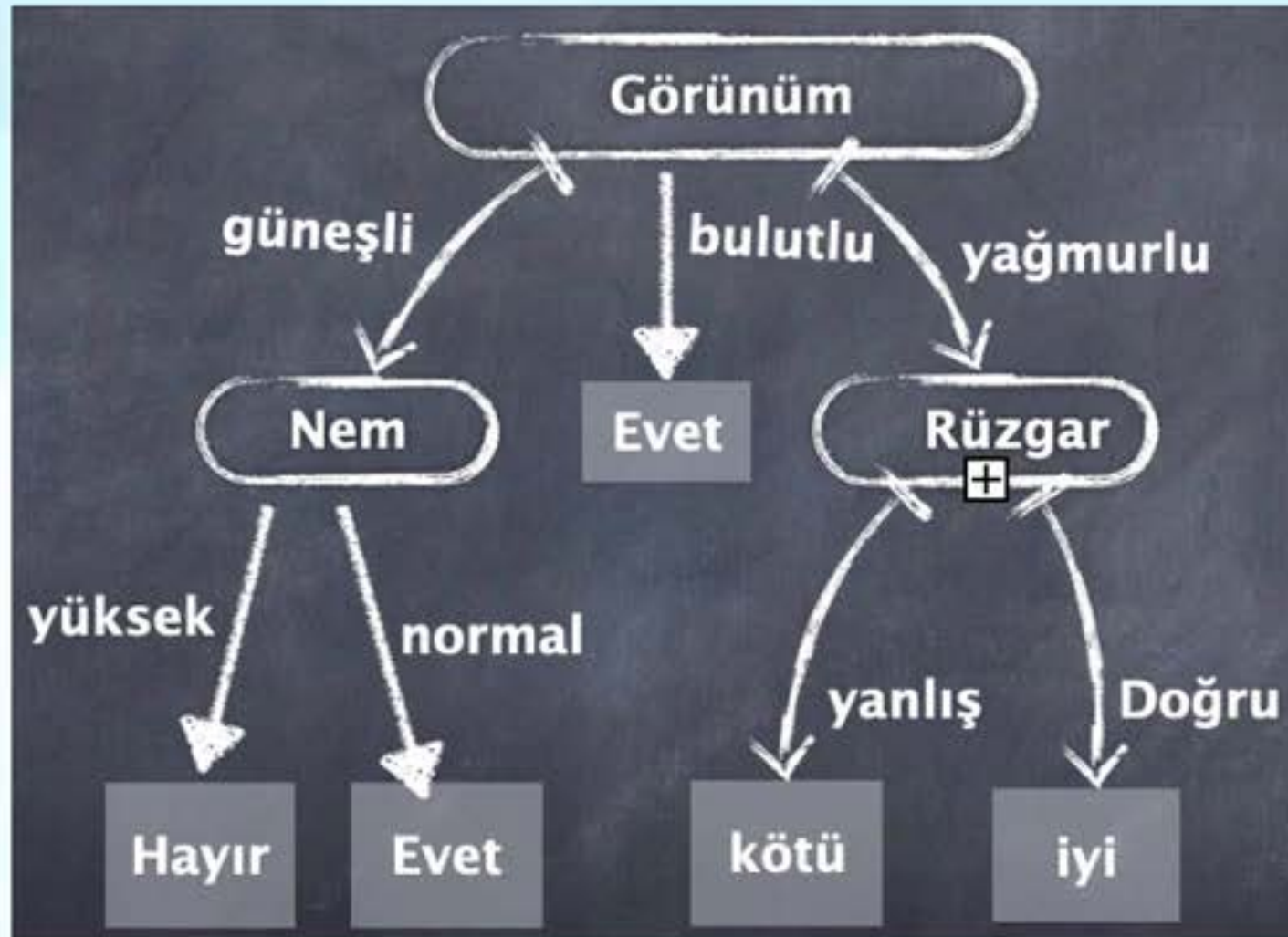
Sonuç olarak, yaprak tarafından atanan sınıftır

Kesin olmayan kurallar üretir

- Hangi sırayla yürütüldükleri önemli değildir
- **Fakat:** ortaya çıkan kurallar gereksiz yere karmaşıktır
- Gereksiz testleri/kuralları kaldırmak için ayıklama gerekir



# Örnek





# Kurallardan Ağaçlara #1

**Daha zor:** kural kümesini ağaca dönüştürme

- Ağaç kurallar arasındaki mahkeme dışı kararı kolayca ifade edemez

**Örnek:** farklı öznelilikleri sınavan kurallar

```
if a ve b then x  
if c ve d then x
```

Simetrisinin kırılması gerekiyor

- Kök düğüm için tek bir test seçmeniz gerekir.
- Karşılık gelen ağaç aynı alt ağaçları içeriyor  
( $\Rightarrow$  "çoğaltılmış alt ağaç sorunu")



# Basit Bir Disjunction için Bir Karar Ağacı

if a ve b then x  
if c ve d then x

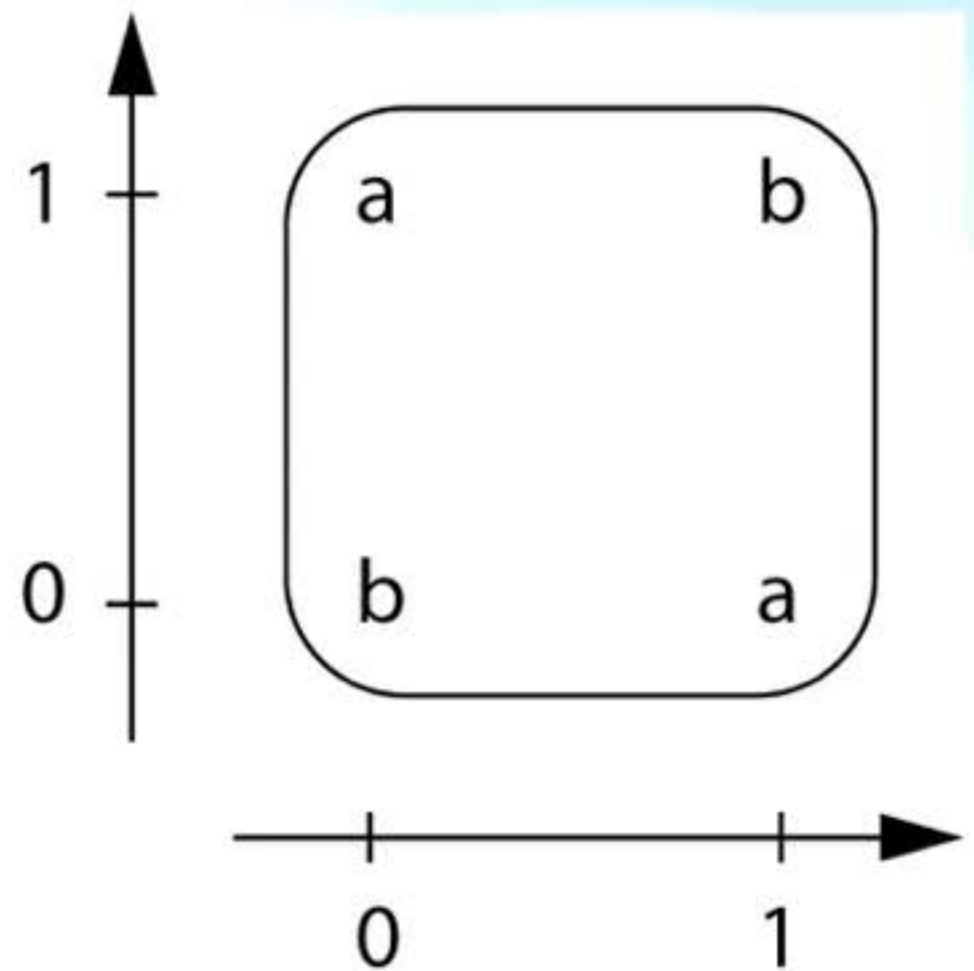


# Kurallardan Ağaçlara #2

Daha zor: ama bazen değil

Örnek: Özel-VEYA (Exclusive-OR) Sorunu

Burada, sınıfın  
yalnızca karşıt  
öznitelik  
değerlerine sahip  
olduğunda "a"  
olmasını isteyin.



# Özel-VEYA (Exclusive-OR) Sorunu

Bu sorun için kurallar nasıl görünür?  
Ağaç neye benzeyecektir?

```
if x = 1 ve y = 0  
    then sınıf = a  
if x = 0 ve y = 1  
    then sınıf = a  
if x = 0 ve y = 0  
    then sınıf = b  
if x = 1 ve y = 1  
    then sınıf = b
```



# Kurallardan Ağaçlara #3

**Daha zor:** kural kümesini ağaca dönüştürme

- Ağaç Varsayılan yan tümceleri kolayca işleyemez

**Örnek:** Her biri 1, 2 veya 3 olabilen dört öznelik.

```
if x = 1 ve y = 1
  then sınıf = a
if z = 1 ve w = 1
  then sınıf = a
aksi takdirde sınıf = b
```

Yeniden çoğaltılmış alt ağaç sorunu.

# Çoğaltılmış Alt Ağacı Olan Karar Ağacı

```
if x = 1 ve y = 1 then sınıf = a  
if z = 1 ve w = 1 then sınıf = a  
aksi takdirde      class = b
```



"Kuralların" popüler olmasının bir nedeni:

Bilginin "Külçeleri (Bilgi kırıntıları)"

Kurallar bağımsız bilgi parçaları mıdır? (Varolan bir kural tabanına kural eklemek kolay görünüyor.)

Bir ağaca eklemek toplam yeniden şekillendirmeye neden olabilir.

**Sorun şudur:** kuralların nasıl yürütülür olduğunu yoksayar

- Kural kümesini yürütmek için iki yolu:
- **Düzenli** kurallar kümesi ("karar listesi")
  - Düzen yorum için önemlidir
- **Düzensiz** kurallar kümesi
- Kurallar çakışabilir ve aynı örnek için farklı sonuçlara yol açabilir



# Kuralları Yorumlama

Ya iki veya daha fazla kural  
çakışıyorsa?

örneğin, farklı kurallar aynı örnek  
için farklı sonuçlara yol açar.

Test örneği için kural geçerli değilse  
ne olur?

Bu, karar ağaçlarında veya karar  
ağaçlarından okunan kurallarda olamaz  
ama oluyor.



**Basitçe:** Bir tür kapalı dünya varsayımı sınıf "boolean" olduğunda ve yalnızca bir sonuç ifade edildiğinde.

**Varsayım:** Örnek "evet" sınıfına ait değilse, "hayır" sınıfına aittir"

**Hile:** Yalnızca "evet" sınıfı için kuralları öğrenin ve "hayır" için varsayılan kuralı kullanın"

```
if x = 1 ve y = 1 then sınıf = a  
if z = 1 ve w = 1 then sınıf = a  
aksi takdirde sınıf = b
```

Kurallar sırası önemli değil. Çakışma yok!

Kural disjunctive normal formda yazılabilir

*i.e.: OR bir sürü AND koşulu.*



# İlişkilendirme Kuralları...

... herhangi bir özniteliği ve öznitelik birleşimini tahmin edebilir

... birlikte bir küme olarak kullanılmak üzere tasarlanmamıştır (birlikte kullanılması amaçlanan sınıflandırma kurallarına karşı)

**Sorun:** çok sayıda olası ilişki

- Çıktının yalnızca en iyi tahmine dayalı ilişkileri gösterecek şekilde kısıtlanması gerekir ==> yalnızca yüksek destek ve yüksek güvene sahip olanları gösterir

**Kapsam** (diğer bir deyişle destek): Kuralın doğru tahmin ettiği örnek sayısı.

**Doğruluk** (diğer bir deyişle güven): Uygulandığı tüm örneklerin bir oranı olarak doğru tahmin ettiği örneklerin sayısı.



Görünüm	Sıcaklık	Nem	Rüzgarlı	Oyun
Güneşli	Sıcak	Yüksek	Yanlış	Hayır
Güneşli	Sıcak	Yüksek	Doğru	Hayır
Bulutlu	Sıcak	Yüksek	Yanlış	Evet
Yağmurlu	Hafif	Yüksek	Yanlış	Evet
Yağmurlu	Serin	Normal	Yanlış	Evet
Yağmurlu	Serin	Normal	Doğru	Hayır
Bulutlu	Serin	Normal	Doğru	Evet
Güneşli	Hafif	Yüksek	Yanlış	Hayır
Güneşli	Serin	Normal	Yanlış	Evet
Yağmurlu	Hafif	Normal	Yanlış	Evet
Güneşli	Hafif	Normal	Doğru	Evet
Bulutlu	Hafif	Yüksek	Doğru	Evet
Bulutlu	Sıcak	Normal	Yanlış	Evet
Yağmurlu	Hafif	Yüksek	Doğru	Hayır

**Örnek:** Normal nem ile 4 serin gün

if sıcaklık = serin then nem = normal

==> Destek = 4, güven = 100%



Minimum destek ve güven eşiğini karşılamayabilir, bu nedenle bu özel kural oluşturulmaz, ancak destek ve güven örneğidir.

Görünüm	Sıcaklık	Nem	Rüzgarlı	Oyun
Güneşli	Sıcak	Yüksek	Yanlış	Hayır
Güneşli	Sıcak	Yüksek	Doğru	Hayır
Bulutlu	Sıcak	Yüksek	Yanlış	Evet
Yağmurlu	Hafif	Yüksek	Yanlış	Evet
Yağmurlu	Serin	Normal	Yanlış	Evet
Yağmurlu	Serin	Normal	Doğru	Hayır
Bulutlu	Serin	Normal	Doğru	Evet
Güneşli	Hafif	Yüksek	Yanlış	Hayır
Güneşli	Serin	Normal	Yanlış	Evet
Yağmurlu	Hafif	Normal	Yanlış	Evet
Güneşli	Hafif	Normal	Doğru	Evet
Bulutlu	Hafif	Yüksek	Doğru	Evet
Bulutlu	Sıcak	Normal	Yanlış	Evet
Yağmurlu	Hafif	Yüksek	Doğru	Hayır

Örnek:

if oyun = evet then rüzgârlı = yanlış

=> Destek = ?, güven = ?



# Bir Kuralın Destegi ve Güveni

**Destek:** doğru tahmin edilen örnek sayısı

**Güven:** kuralın uygulandığı tüm örneklerin oranı olarak doğru tahminlerin sayısı

**Örnek:** Normal nem ile 4 serin gün

```
if sıcaklık = serin then nem = normal
```

=> Destek = 4, güven = 100%

**Normalde:** önceden belirtilen minimum destek ve güven (örneğin, hava durumu verileri için 58 kural destek  $\geq 2$  ve güven  $\geq 95\%$  içerir)



# İlişkilendirme

## Kurallarını Yorumlama

Görünüm	Sıcaklık	Nem	Rüzgarlı	Oyun
Güneşli	Sıcak	Yüksek	Yanlış	Hayır
Güneşli	Sıcak	Yüksek	Doğru	Hayır
Bulutlu	Sıcak	Yüksek	Yanlış	Evet
Yağmurlu	Hafif	Yüksek	Yanlış	Evet
Yağmurlu	Serin	Normal	Yanlış	Evet
Yağmurlu	Serin	Normal	Doğru	Hayır
Bulutlu	Serin	Normal	Doğru	Evet
Güneşli	Hafif	Yüksek	Yanlış	Hayır
Güneşli	Serin	Normal	Yanlış	Evet
Yağmurlu	Hafif	Normal	Yanlış	Evet
Güneşli	Hafif	Normal	Doğru	Evet
Bulutlu	Hafif	Yüksek	Doğru	Evet
Bulutlu	Sıcak	Normal	Yanlış	Evet
Yağmurlu	Hafif	Yüksek	Doğru	Hayır

Yorum açık değil

```
if rüzgârlı = yanlış ve oyunOynama = Hayır  
then görünüm = güneşli ve nem = yüksek
```

Bu bir kısayol değildir

```
if rüzgârlı = yanlış ve oyunOynama = Hayır  
then görünüm = güneşli  
if rüzgârlı = yanlış ve oyunOynama = Hayır  
then nem = yüksek
```

- Ancak 1. kural, aşağıdakilerin geçerli olduğu anlamına gelir:

```
if nem = yüksek ve rüzgârlı = yanlış ve oyunOynama = Hayır  
then görünüm = güneşli
```