

# Veri Madenciliği

Güz 2023

Ders 7

- Çıktı: Ağaçlar
- WEKA girişi

# Ağaçları Sayısal Tahmin İçin De Kullanabilir miyiz?

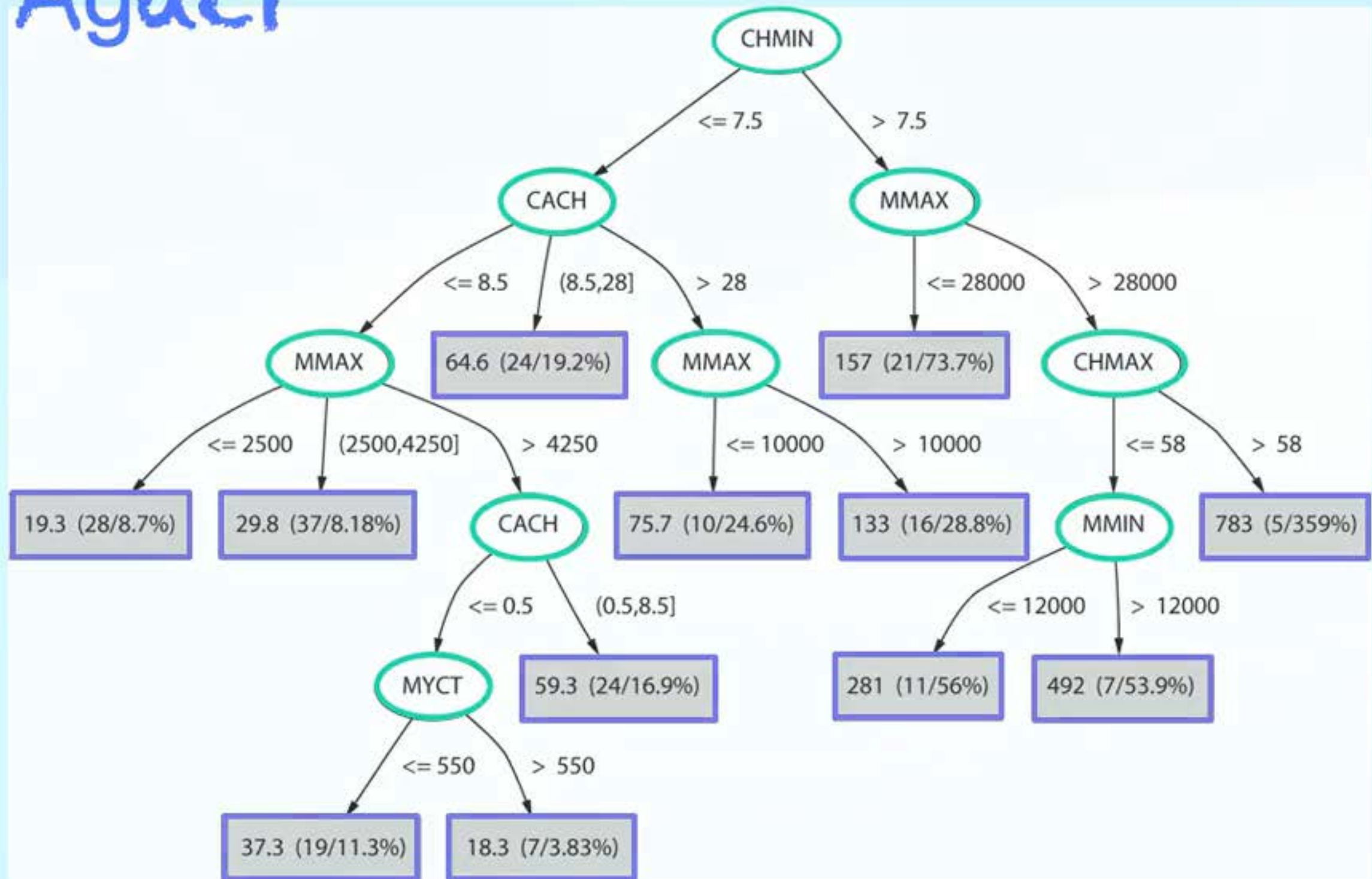
- Regresyon: sayısal niceliği tahmin eden bir ifadeyi hesaplama işlemi
- Regresyon ağacı: Her yaprağın sayısal bir miktarı öngördüğü "**karar ağacı**"
  - Tahmin edilen değer, yaprağa ulaşan eğitim örneklerinin ortalama değeridir
- Model ağacı: Yaprak düğümlerinde doğrusal regresyon modelleriyle "**regresyon ağacı**"
  - Doğrusal yamalar sürekli fonksiyona yaklaştırır.



# CPU Verileri için Doğrusal Regresyon

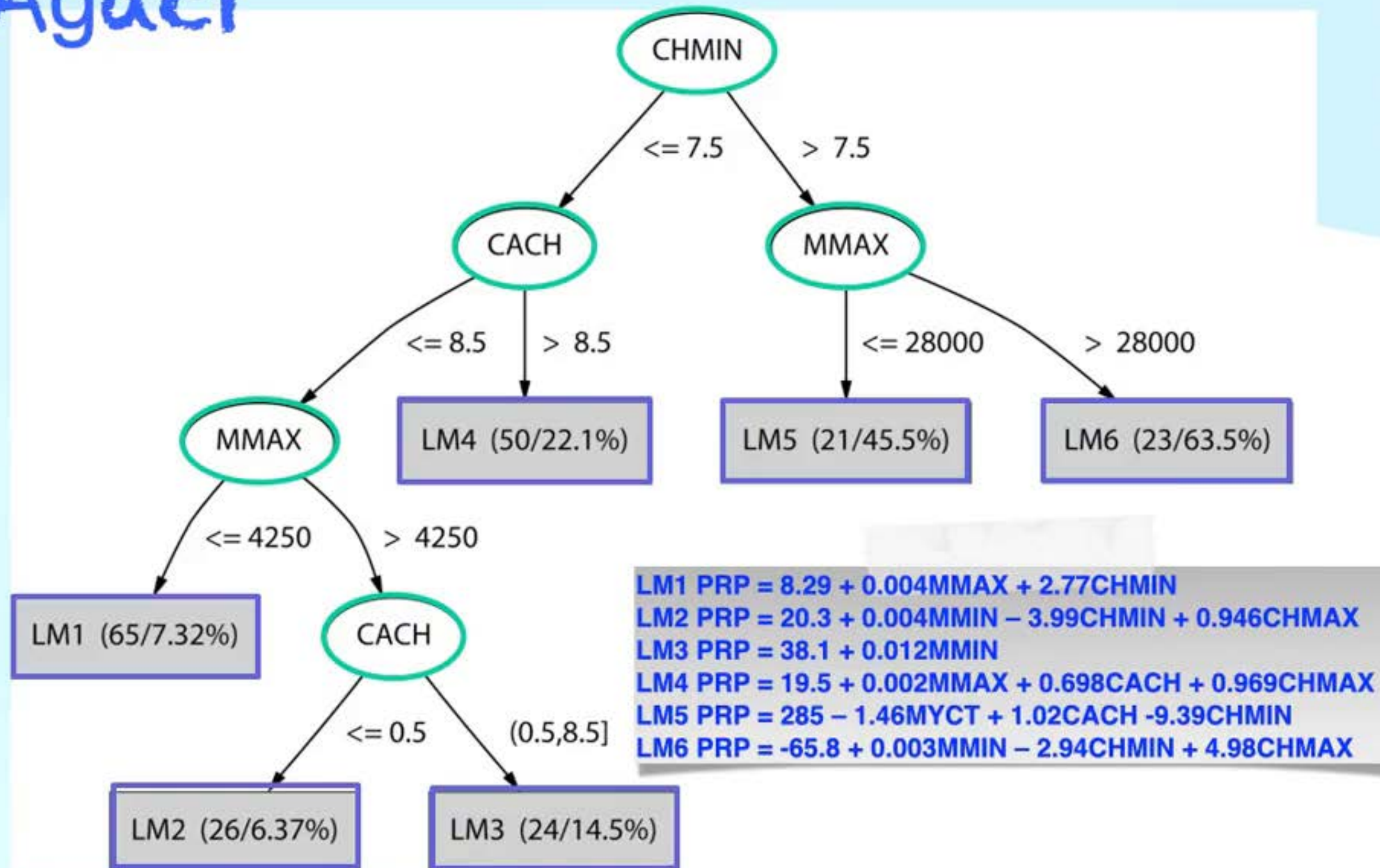
$$\begin{aligned} \text{PRP} = & \\ & -56.1 \\ & + 0.049 \text{ MYCT} \\ & + 0.015 \text{ MMIN} \\ & + 0.006 \text{ MMAX} \\ & + 0.630 \text{ CACH} \\ & - 0.270 \text{ CHMIN} \\ & + 1.46 \text{ CHMAX} \end{aligned}$$

# CPU Verileri için Regresyon Ağacı





# CPU Verileri için Model Ağacı



# WEKA

## Bilgi Analizi için Waikato Ortamı

Bir Windows, Linux veya Mac bilgisayara indirmek için:

<https://waikato.github.io/weka-wiki/downloading-weka/>



# WEKA'ya Giriş

- birçok veri madenciliği ve makine öğrenimi algoritmasının açık kaynak kodu da dahil olmak üzere
  - veriler üzerinde ön işleme
  - sınıflandırma
  - kümeleme
  - birliktelik kuralı çıkarmayı da içeren koleksiyonudur.
- Yeni Zelanda'daki Waikato Üniversitesi'ndeki araştırmacılar tarafından oluşturuldu.
- Java tabanlı (ayrıca açık kaynak).

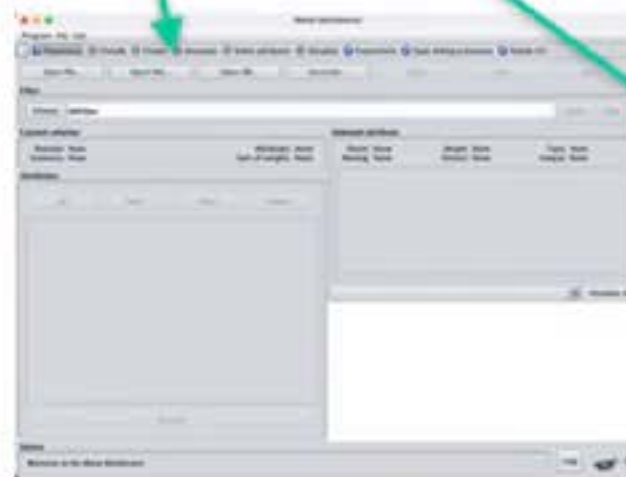
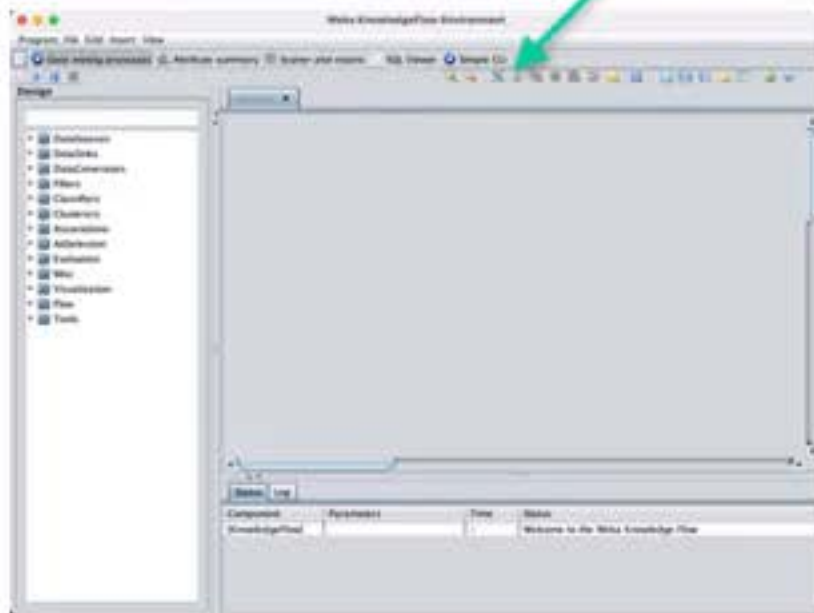
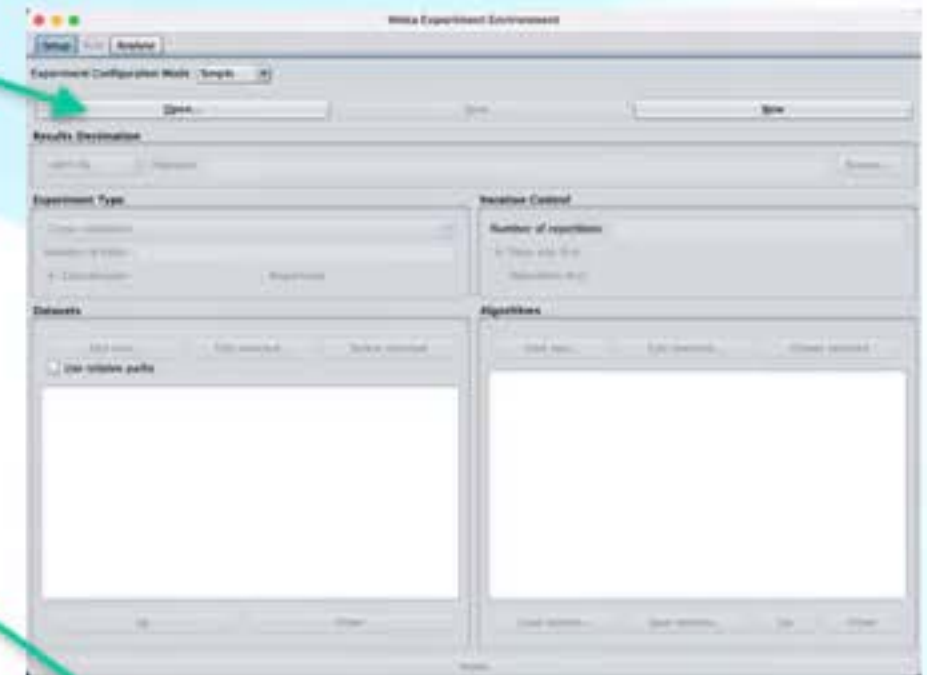
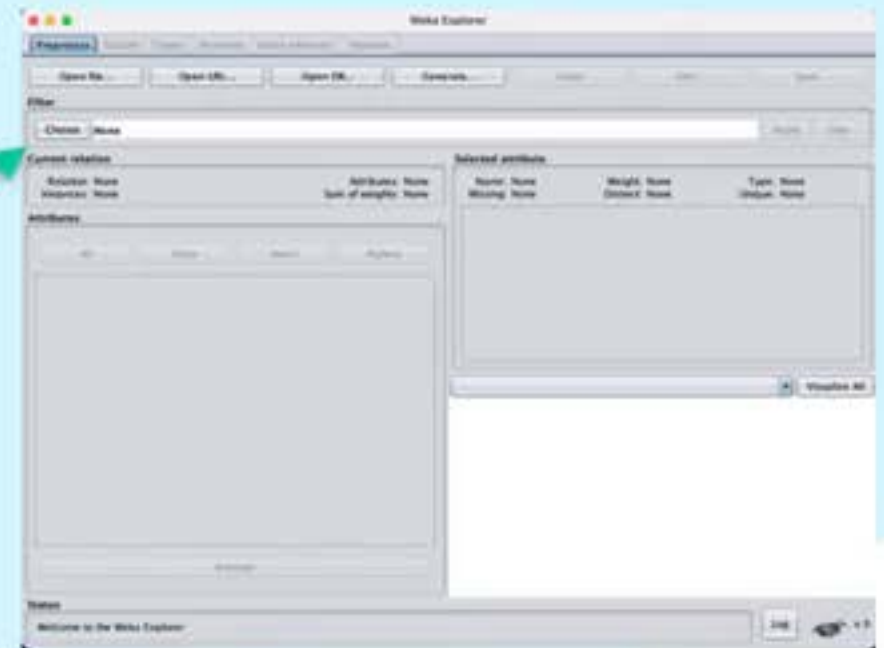
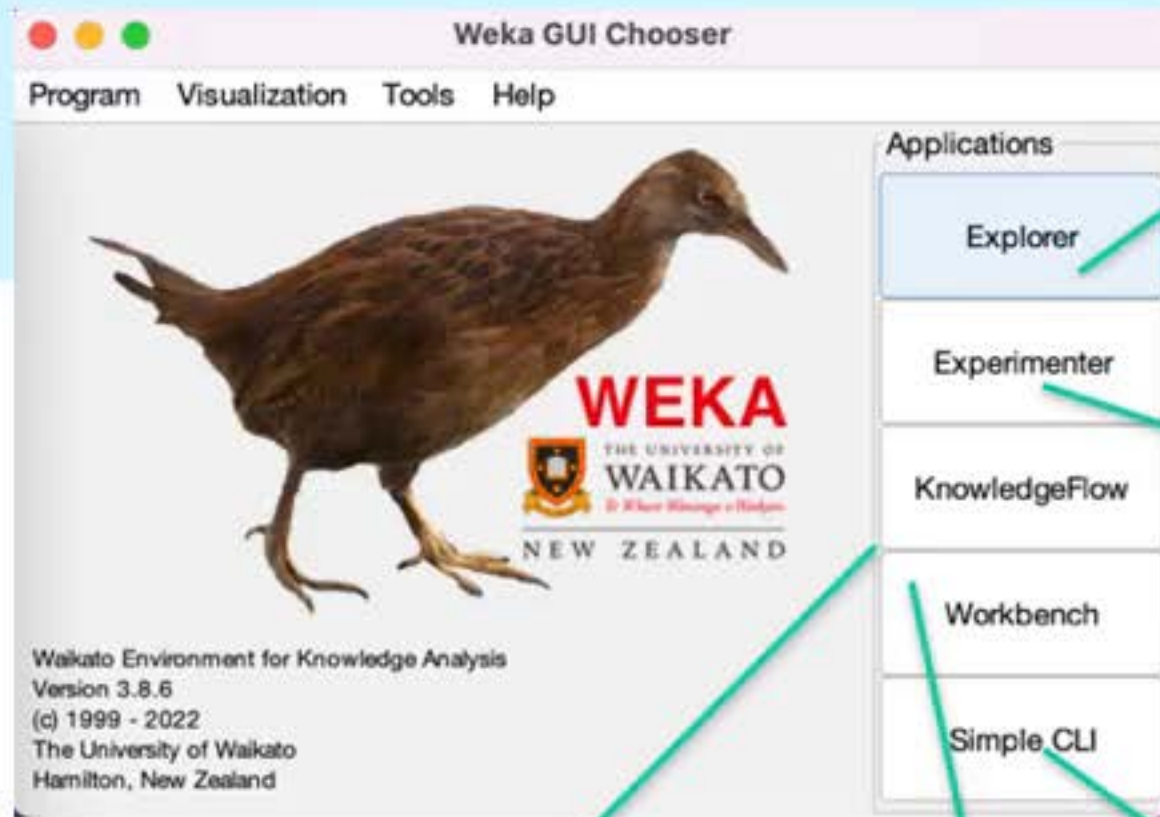


# WEKA'nın Ana Özellikleri

- ~ 49 veri ön işleme aracı
- ~ 76 sınıflandırma/regresyon algoritması
- ~ 8 kümeleme algoritması
- ~15 özellik/alt küme değerlendiricisi + özellik seçimi için 10 arama algoritması
- ~ Birliktelik kurallarını bulmak için 3 algoritma
- 3 grafik kullanıcı arayüzü
  - ▶ "The Explorer" (keşif amaçlı veri analizi)
  - ▶ "Experimenter" (deneysel ortam)
  - ▶ "KnowledgeFlow" (yeni süreç modelinden ilham alan arayüz)



# WEKA



# WEKA Uygulama Arayüzü

- Gezgin
  - ön işleme, nitelik seçimi, öğrenme, görselleştirme
- Deneyci
  - makine öğrenimi algoritmalarını test etme ve değerlendirme
- Bilgi Akışı
  - KDD (Bilgi Keşfi /Veriden/Veritabanlarından/Veri madenciliği ile) sürecinin görsel tasarımı
- WEKA Tezgahı
  - Ön işleme, filtreleme, verinin testi, ...
- Basit Komut satırı
  - Komut yazmak için basit bir arayüz



# WEKA Fonksiyonları ve Araçları

- Ön İşleme Filtreleri
- Öznitelik seçimi
- Sınıflandırma/Regresyon
- Kümeleme
- İlişkilendirme keşfi
- Görselleştirme

# WEKA: Artıları ve Eksileri

## • Artıları

- Açık kaynak,
  - Özgür
  - genişletilebilir
  - Diğer java paketlerine entegre edilebilir
- GUI (Grafik Kullanıcı Arayüzleri)
  - Nispeten kullanımı kolay
- Özellikleri
  - Bireysel deneme geliştirin veya
  - KDD aşamaları oluşturun

## • Eksileri

- Uygun ve yeterli dokümantasyon eksikliği
- Sistemler sürekli güncellenmektedir (Mutfak Lavabo Sendromu)