



# Popüler Veri Madenciliđi Araçları

Dr.Günay TEMÜR

## Veri madenciliđi,

- Eldeki verilerden üstü kapalı, çok net olmayan, önceden bilinmeyen ancak potansiyel olarak kullanışlı bilginin çıkarılmasıdır. Bu da; **kümeleme, veri özetleme, deđişikliklerin analizi, sapmaların tespiti** gibi belirli sayıda teknik yaklaşımları içerir.
- **Veri madenciliđi** konusunda bahsi geçen geniş verideki geniş kelimesi, tek bir iş istasyonunun belleđine sığamayacak kadar büyük veri kümelerini ifade etmektedir. Yüksek hacimli veri ise, tek bir iş istasyonundaki ya da bir grup iş istasyonundaki disklere sığamayacak kadar fazla veri anlamındadır. Dađıtık veri ise, farklı cođrafi konumlarda bulunan verileri anlatır.

# Popüler Araçlar

- *Bu aşamada veriyi işleyebilmek için kullanılan popüler araçlardan bazıları sizler için paylaşılacaktır.*

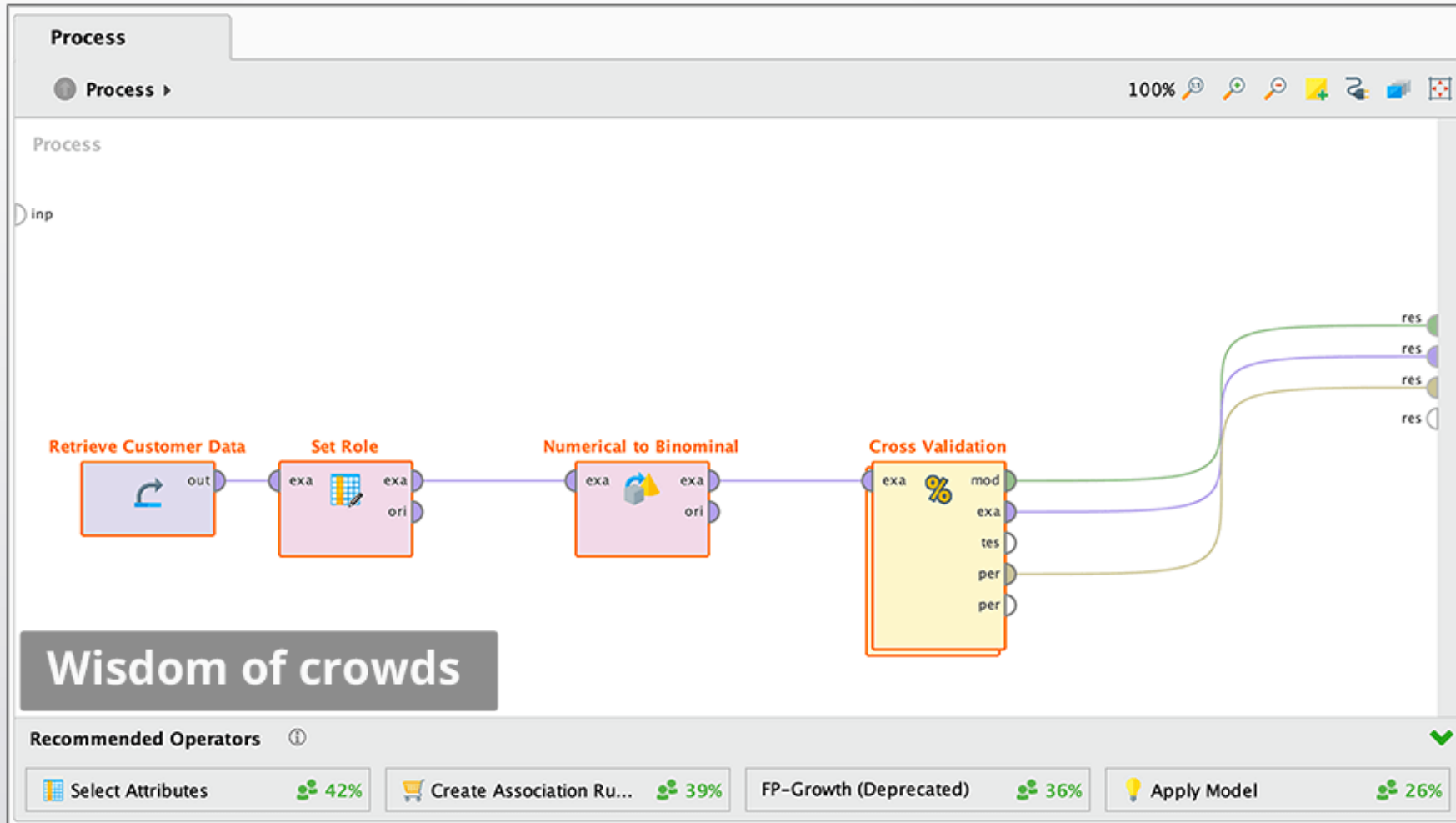
# 1) rapidminer

- Java programlama dili kullanılarak yazılmıştır. Kod yazma ihtiyacını neredeyse sıfıra indirdiği için kullanıcılar tarafından tercih edilir, kod ihtiyacının az olması en temelde hata riskini minimuma indirir. **RapidMiner**, veri yükleme ve dönüşüm (çıkarma, dönüşüm, yükleme (ETL)), veri ön işleme ve görselleştirme, öngörülü analitik ve istatistiksel modelleme, değerlendirme ve dağıtım da dahil olmak üzere veri madenciliği ve **deep learning** özelliklerine sahiptir.

# 1) rapidminer

- **RapidMiner**, analitik iş akışlarını tasarlamak ve yürütmek için bir GUI sağlar. Bu iş akışlarına RapidMiner'de "Süreçler" denir ve bunlar birden fazla "Operatör" içerir. Her operatör süreç içinde tek bir görev gerçekleştirir ve her operatörün çıktısı bir sonraki işin girişini oluşturur. Alternatif olarak, motor diğer programlardan çağrılabilir veya API olarak kullanılabilir. Bireysel fonksiyonlar komut satırından çağrılabilir. RapidMiner öğrenme şemaları, modeller ve algoritmalar sağlar ve R ve Python komut dosyalarını kullanarak genişletilebilir.
- RapidMiner işlevselliği, RapidMiner Pazaryeri üzerinden sunulan ek eklentilerle genişletilebilir. RapidMiner Marketplace, geliştiricilerin veri analizi algoritmaları oluşturması ve bunları topluluğa yayınlaması için bir platform sağlar.

1)





## 2) Weka

- Birkaç standart veri madenciliği görevini, daha spesifik olarak, veri ön işleme, kümeleme, sınıflandırma, gerileme, görselleştirme ve özellik seçimi konularını desteklemektedir. **Weka**, veri analizleri ve tahmini modelleme için görselleştirme araçları ve algoritmaları topluluğu ile birlikte bu işlemlere kolay erişim için grafik kullanıcı arayüzleri içerir. Grafik kullanıcı arayüzleri sayesinde kullanım kolaylığı sağlar.



## 2) Weka

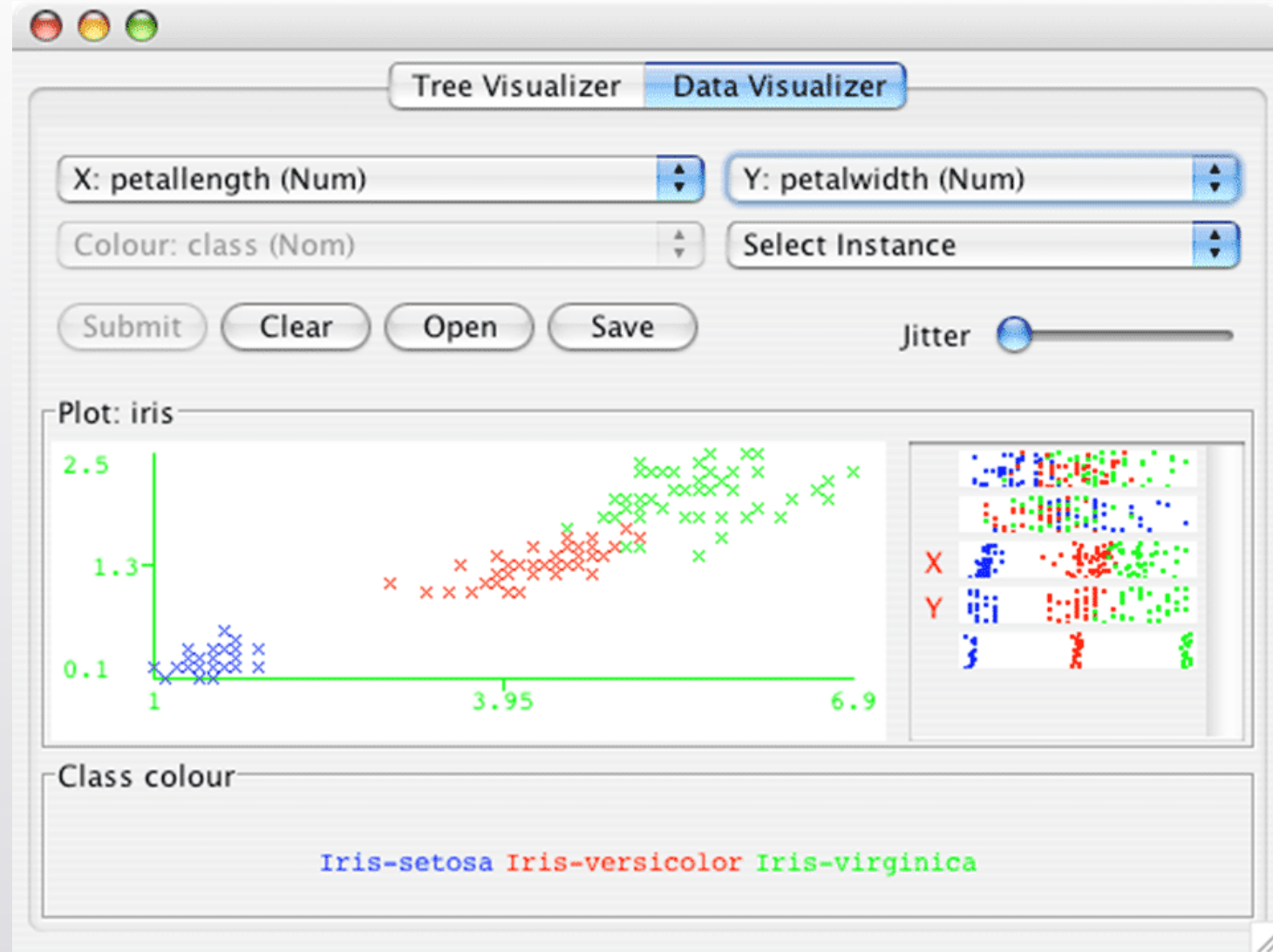
- **Weka, Java Database Connectivity** kullanarak SQL veritabanlarına erişim sağlar ve bir veritabanı sorgusu tarafından döndürülen sonucu işleyebilir. Weka Deeplearning4j ile derin öğrenime erişim sağlar. Çok ilişkili veri madenciliği yapma yeteneği yoktur, ancak bağlı veritabanı tablolarından oluşan bir database'i , Weka kullanılarak işlenmeye elverişli tek bir tabloya dönüştürmek için ayrı bir yazılım bulunur. Sıra modellemesi özelliğine sahip olması bir diğer avantajıdır.



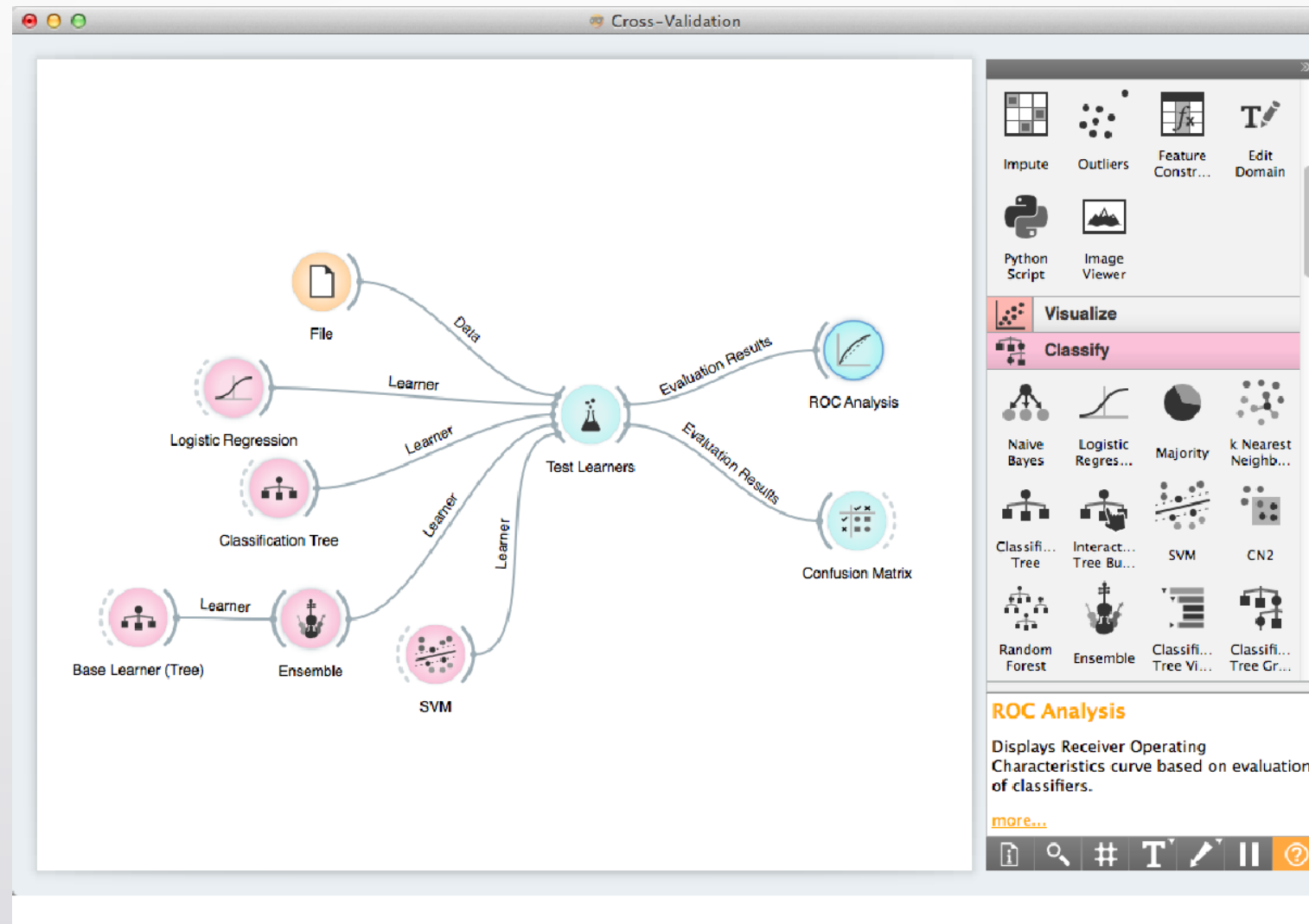
## 2) Weka



WEKA



### 3) Orange



### 3) Orange



- Veri görselleştirme, makine öğrenmesi, veri madenciliği ve veri analizi için bileşen tabanlı bir görsel programlama yazılımı paketidir.
- Orange bileşenlerine pencere öğeleri denir ve basit veri görselleştirme, alt grup seçimi ve önişleme, öğrenme algoritmalarının ampirik değerlendirmesine ve tahmini modellemeye kadar uzanır.

### 3) Orange



- Görsel programlama, iş akışlarının önceden tanımlanmış veya kullanıcı tarafından tasarlanmış widget'ları bağlayarak oluşturulduğu bir arayüz üzerinden gerçekleştirilirken, ileri düzey kullanıcılar Orange'ı veri işleme ve ya bir Python kütüphanesi olarak kullanabilirler.
- Turuncu, kullanıcının widget yerleştirdiği ve bir veri analizi iş akışı oluşturduğu bir tuval arayüzünden oluşur. Pencere öğeleri, verileri okurken, veri tablosu gösterirken, özelliklerin seçilmesi, öğrenme tahmincileri, öğrenme algoritmalarının karşılaştırılması, veri öğelerinin görselleştirilmesi vb. Gibi temel işlevleri sağlar. Kullanıcı, görselleştirmeleri interaktif olarak keşfedebilir veya seçilen alt kümeyi diğer widget'lere besleyebilir.

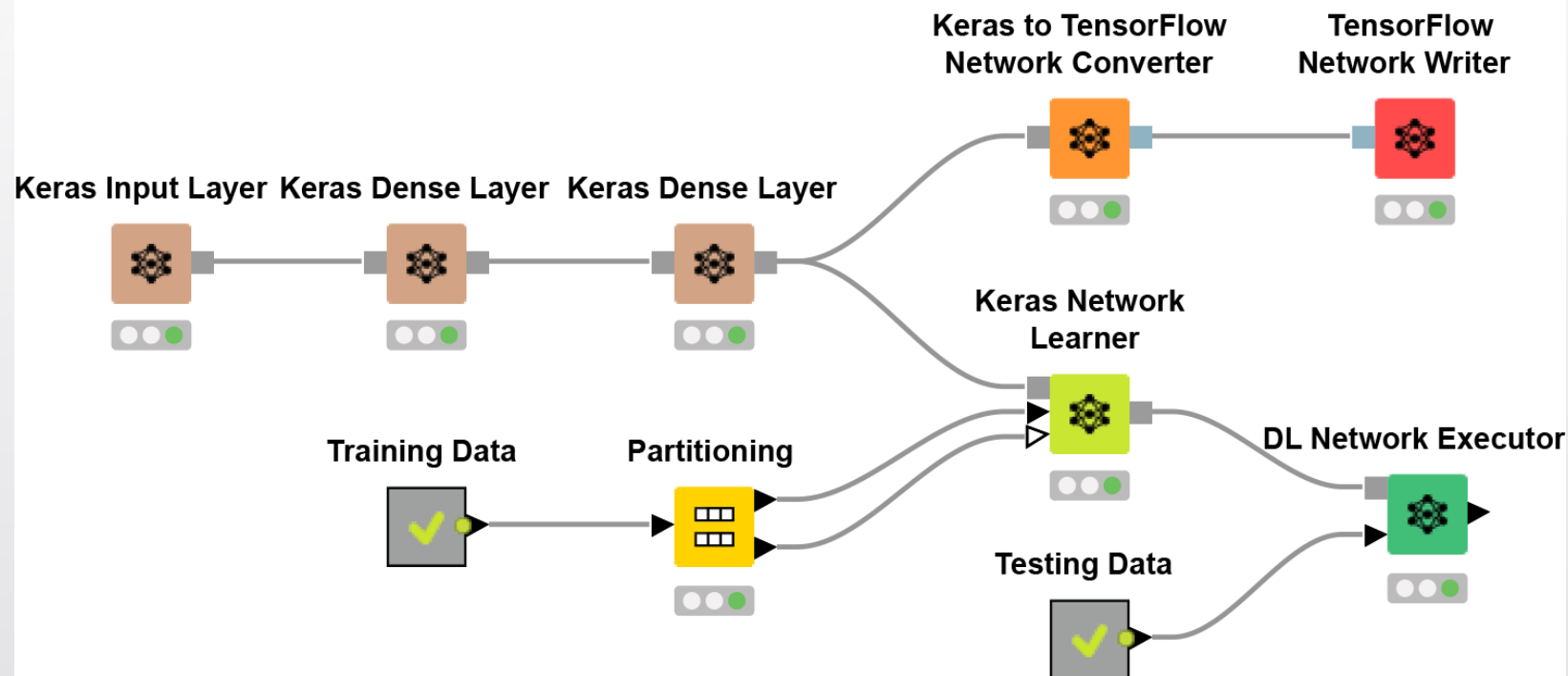
## 4) Knime

- Veri ön işleme üç ana bileşene sahiptir: ekstraksiyon, dönüşüm ve yükleme. KNIME her üçünü de yapıyor. Veri işleme için düğümlerin birleştirilmesine izin vermek için size bir grafik kullanıcı ara birimi sağlar. Açık kaynak veri analizi, raporlama ve entegrasyon platformudur. KNIME ayrıca modüler veri boru hattı konsepti aracılığıyla makine öğrenimi ve veri madenciliği için çeşitli bileşenleri entegre eder ve iş zekası ve finansal veri analizinin önünü açar.

## 4) Knime

- Java'da yazılmış ve Eclipse üzerine kurulu olan KNIME, genişletmek ve eklentileri eklemek kolaydır. İlave işlevler hareket halindeyken eklenebilir. Çekirdek sürümde birçok veri entegrasyon modülü bulunmaktadır.

## 4) Knime



# Diğer Programlar

- **4) R**
  - İstatistiksel hesaplama ve grafikler için bir programlama dili ve özgür yazılım ortamıdır.
- **6-Rattle**
  - R kullanarak sağlanan veri madenciliği için bir grafik kullanıcı ara birimidir.
- **7- Oracle**
  - Veri tabanında çeşitli veri madenciliği algoritmaları uygulanır. Bu uygulamalar doğrudan Oracle database çekirdeği ile bütünleşir ve doğal olarak ilişkisel veritabanı tablolarında depolanan veriler üzerinde çalışırlar.



# Diğer Programlar

- **8-Tanagra**

- Araştırma ve akademik amaçlarla Ricco Rakotomala tarafından Lumière Üniversitesi Lyon 2, Fransa'da geliştirilen, makine öğrenme yazılımının ücretsiz bir paketidir.

- **9- Sas**

- SAS, çeşitli kaynaklardan veri toplama, değiştirme, yönetme ve alma ve bunun üzerine istatistiksel analiz yapabilen bir yazılım paketidir.

- **10- Spss**

- IBM SPSS Modeler, IBM'den veri madenciliği ve metin analiz yazılımı uygulamasıdır.