

**Estudiante(s):** Marco Antonio Ayala Lituma

**Fecha de entrega:** 5 de mayo del 2021

**Tema:** Comparativa de herramientas para Minería de Base de datos

**Paralelo:** A

## TAREA SEMANA 5

# Resumen

- Características
- Plataformas soportadas
- Ventajas y Desventajas
- Conclusiones
- Referencias

# Desarrollo

## Introducción

Como data mining, también conocida como minería de datos, se describen los procedimientos algorítmicos para la evaluación de datos aplicados a bloques de datos con un tamaño y una complejidad determinados. Su función es la de extraer la información oculta en grandes volúmenes de datos, especialmente en las masas de información conocidas como big data, y además reconocer tendencias, relaciones y patrones ocultos en ellas. Para que este proceso se pueda llevar a cabo se recurre a las data mining tools.

Con el término data mining no se hace referencia ni al proceso de generación de datos ni a los bloques de datos en sí, sino a su análisis. Asimismo, el data mining tampoco es pura estadística a pesar de que muchos de los procedimientos utilizados provienen de ella, sino un procedimiento interdisciplinario que combina los conocimientos de informática y matemáticas con la tecnología del machine learning (especialmente el aprendizaje no supervisado) y de la inteligencia artificial. Estos métodos de alto rendimiento se integran en un solo software que permite la evaluación de la mayor cantidad de datos posible.

## Características

A continuación, analizamos y comparamos las mejores herramientas de data mining del mercado hoy en día: RapidMiner, WEKA y Orange. Es bien conocido que los usuarios usan más de una, combinándolas entre sí, pues tienen puntos fuertes diferentes. No obstante, si es una de las primeras veces que recurre a este tipo de programas, también puedes conseguir grandes avances con un único software polivalente.

## RapidMiner

RapidMiner, antes conocida como YALE, siglas de “Yet Another Learning Environment”, es un data mining software muy conocido. De acuerdo con una encuesta de KDnuggets realizada en el año 2014, esta tool se constituía como la herramienta de data mining más usada. Destaca por permitir el acceso gratuito y por su fácil manejo dado que no requiere un conocimiento elaborado en programación, sin olvidar la gran selección de operadores que ofrece. Especialmente son las startups las que recurren a ella.

RapidMiner está escrita en Java y contiene más de 500 operadores con diferentes enfoques para mostrar las conexiones en los datos: hay opciones para data mining, text mining o web mining, pero también análisis de sentimiento o minería de opinión. Asimismo, el programa puede importar tablas Excel, archivos SPSS y masas de datos de diferentes bases de datos e integra los programas de data mining WEKA y R. Todo ello pone de relieve el carácter polivalente de este software.

RapidMiner participa en todos y cada uno de los pasos del proceso de data mining, interviniendo también en la visualización de los resultados. La herramienta está formada por tres grandes módulos: RapidMiner Studio, RapidMiner Server y RapidMiner Radoop, cada uno encargado de una técnica diferente de minería de datos. Asimismo, RapidMiner prepara los datos antes del análisis y los optimiza para su rápido procesamiento. Para cada uno de estos tres módulos hay una versión gratuita y diferentes opciones de pago.

El punto fuerte de RapidMiner, si se compara con el resto de software de data mining, reside en los análisis predictivos, es decir, en la previsión de desarrollos futuros basándose en los datos recopilados.

### **WEKA**

WEKA (Waikato Environment for Knowledge Analysis) es un software de código abierto desarrollado por la Universidad de Waikato en la primera mitad de los años noventa. Basada en Java y compatible con Windows, macOS y Linux, la interfaz gráfica de usuario facilita el acceso a este software que también ofrece conexión a bases de datos SQL, siendo capaz de procesar en ellas los datos solicitados. Asimismo, presenta un sinnúmero de funciones de aprendizaje automático y segunda tareas tan relevantes del data mining como el análisis de clústeres, de correlación o de regresión, así como la clasificación de datos, punto fuerte este último del data mining software al usar redes de neuronas artificiales, árboles de decisión y algoritmos ID3 o C4.5. No obstante, este programa no es tan potente en aspectos como el análisis de clústeres, en el que solo se ofrecen los procedimientos más importantes.

Otra desventaja es que este software presenta problemas de procesamiento cuando hay que tratar grandes cantidades de datos, dado que intenta cargar el programa de data mining completo en la memoria de trabajo. WEKA ofrece para ello como solución una línea de comandos sencilla (CTL) que aligera el tráfico de un gran volumen de datos.

### **Orange**

El software de data mining Orange existe desde hace más de 20 años como proyecto de la Universidad de Liubliana. El núcleo del software se escribió en C++, aunque poco después se amplió el programa al lenguaje de programación Python que solo se usa como lenguaje de entrada. Las operaciones más complejas, sin embargo, se llevan a cabo en C++. Orange

es un software muy extenso que demuestra todo lo que se puede conseguir con Python, ya que ofrece aplicaciones de gran utilidad para el análisis de datos y de texto así como características de aprendizaje automático.

Además, en el ámbito del data mining trabaja con operadores para la clasificación, regresión y clustering e integra una programación visual. De hecho, es muy llamativo que los usuarios destaquen lo entretenido que es usar esta herramienta en comparación con otras: tanto si se empieza con la minería de datos o se es más experto, a todos los usuarios les fascina Orange. Esto se debe a que, por un lado, ofrece un sistema de visualización de datos atractivo para trabajar y, por otro, alcanza esta visualización con rapidez y facilidad. El programa prepara los datos de forma visual, convirtiendo la comprensión de gráficas así como el procesamiento de análisis de datos en tareas muy sencillas, lo que a su vez facilita a los usuarios tomar decisiones rápidamente en el ámbito profesional.

Otra ventaja para los menos versados: existe un sinnúmero de tutoriales sobre la herramienta. Una particularidad de Orange es que, además, va aprendiendo las preferencias de sus usuarios y se comporta en función de ellas, lo que simplifica enormemente el proceso de data mining para el usuario.

#### Plataformas soportadas

Sistema Operativo	RapidMiner	WEKA	Orange
Windows	X	X	X
Linux	X	X	X
macOS	X	X	X

#### Ventajas y Desventajas

Sistema Operativo	Ventajas	Desventajas
Rapid Miner	Herramienta completa con diversos módulos de analista de datos, más popular.	Consume muchos recursos del computador. Dificultad de aprendizaje.
WEKA	Buena documentación, orientado a pymes y análisis de datos no tan complejos.	No es muy potente a la hora de trabajar con muchos orígenes de datos
Orange	Esta herramienta es fácil, de utilizar no tan pesada y muy potente para usuarios novatos Herramienta visual muy versátil	X

#### Conclusiones

Las herramientas de minería de base de datos son tecnologías que nos facilitan en la capacidad de capturar, gestionar y procesar de forma veraz todo tipo de datos, utilizando dichas herramientas o softwares que identifican patrones comunes. Estos patrones podrían ser características específicas de los consumidores, generación de parámetros, métricas, entre muchos otros. Y, tienen la capacidad de cambiar la manera de hacer negocios, ya que permiten aumentar la rentabilidad y productividad de las compañías.

A la hora de elegir cualquier herramienta se debe analizar las que permita escalabilidad y rendimientos a la hora de generar procesos de análisis de información, eso ya queda a cargo del analista de base de datos y con los gestores de base de datos que se este trabajando y recursos externos para una correcta implementación de datos internos y externos.

## Referencias

Data Mining recuperado en: [Cinco herramientas de Data Mining o minería de datos \(culturacrm.com\)](http://culturacrm.com)

Software de DataMinig, recuperado en: [Software de data mining: las mejores herramientas - IONOS](#)