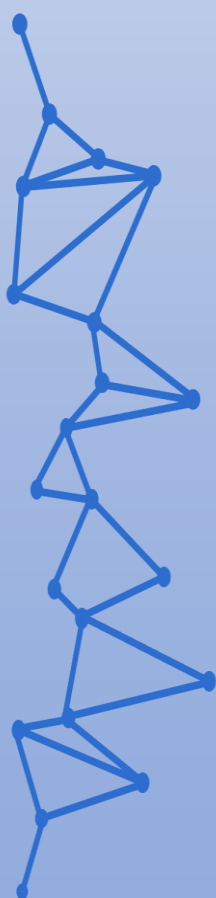




## Curso de Especialización de Inteligencia Artificial y Big Data (IABD)



### Sistemas de aprendizaje Automático

UD04. Casos prácticos de aplicación.  
Resumen.

JUAN ANTONIO GARCIA MUELAS

---

**BigML** es una plataforma online que **permite plantear y resolver problemas de aprendizaje automático** (supervisado y no supervisado) con los principales algoritmos que conocemos de esta técnica de Inteligencia Artificial.

Cuenta con una **versión gratuita** en la que podemos trabajar con una **base de datos** de hasta **16MB**.

Puedes crear tu cuenta con tu usuario Google, GitHub o Amazon.

Dentro de las categorías de un dataset podemos previsualizar antes de adquirirlo datos como su nombre, tipología, cantidad, datos perdidos, datos erróneos y un histograma de distribución.

Si encuentro un dataset en la plataforma BigML que me gusta, puedo descargar la base de datos correspondiente en formato .csv.

También puedes trabajar en BigML con **bases de datos propias** (aportadas por ti), cargando el archivo de dicha base de datos en el apartado “**Source**” y luego generando el dataset correspondiente.

Cuando trabajo **con mis propias bases de datos**, si quiero prescindir de alguna categoría o descartar instancias incompletas o erróneas debo gestionarlo **después de** haber **cargado** el dataset.

Puedes analizar los datos con los que se va a trabajar antes de hacer un entrenamiento de Aprendizaje Automático, aunque no es la función principal de la plataforma.

La vista **Scatterplot** de los datasets en BigML sirve para **repasar relaciones** entre las diferentes categorías de nuestro dataset, confrontándolas **de dos en dos y ver** las **gráficas** que generan.

Podemos seleccionar la **categoría** como **campo objetivo** para realizar un entrenamiento.

Es **posible editar** el nombre de las **categorías** de un dataset antes de lanzar el entrenamiento.

Podemos utilizar el mismo dataset para **hacer entrenamientos con algoritmos diferentes** y, de hecho, es habitual en muchos casos para comparar niveles de exactitud y precisión alcanzados, para quedarse con el mejor modelo.

Siempre que hagamos un modelo de Inteligencia Artificial (tanto con BigML como con cualquier otra plataforma) es conveniente reservar un 20% de los datos sin usarlos en la fase de entrenamiento para poder hacer posteriormente una evaluación realista del modelo, aunque podemos definir la cantidad que reservamos para cada parte.

En BigML puedes iniciar entrenamientos de dos formas: De manera rápida (usando parámetros genéricos por la propia plataforma) o definiendo tú los parámetros disponibles para cada algoritmo.

A mayor número de instancias y categorías en un dataset mayor es el tiempo de cálculo en el entrenamiento.

En BigML al terminar un entrenamiento de aprendizaje automático ves diferentes representaciones gráficas del modelo dependiendo del algoritmo que hayas utilizado en dicho entrenamiento.

Para poder hacer la evaluación de un modelo ya entrenado necesitamos una base de datos que incluya también el campo objetivo en el entrenamiento, para que la herramienta pueda comparar caso a caso el resultado real con el pronosticado por el modelo.

La **vista “confidence”** que se puede **visualizar** en un modelo ya entrenado con el algoritmo de Árbol de Decisión representa: **La confiabilidad** del modelo.

Una vez terminado un entrenamiento supervisado con algoritmo de Árbol de Decisión podemos hacer predicciones:

- ✓ Introduciendo dato a dato un nuevo caso.
- ✓ Cargando a través de una base de dato múltiples casos a la vez.
- ✓ Introduciendo a la vez todos los datos de una nueva instancia.

La **vista “prediction”** que se puede **visualizar** en un modelo ya entrenado con el algoritmo de Árbol de Decisión representa: Los **valores del campo objetivo** que ha encontrado para cada caso.

En un entrenamiento de tipo “no supervisado” en BigML realizado con el algoritmo “cluster” podemos decidir hacerlo con K-Means (eligiendo nosotros el número de clústers) o con G-Means (dejando al propio modelo que decida cuántos clusters definir) .

Estos modelos entrenados con el algoritmo “cluster”, los representa gráficamente con círculos.