

# Modelar sobre plataforma de ML

## Programa Ejecutivo IA & Deep Learning

**Profesor:**

**Andrés González García** - CTO y Socio Cofundador de CleverData.io  
agonzalez@cleverdata.io

# Índice

- Recursos recomendados.
- Ejercicio 1. Ver un vídeo y contestar a las preguntas.

# Recursos recomendados

## Artículos

- [Una introducción visual al ML](#). Explicación muy visual y didáctica de cómo funcionan los árboles de decisión. La mejor que he visto nunca.
- [Model Tuning and the Bias-Variance Tradeoff](#). Segunda parte del artículo anterior que explica dos de los problemas clásicos de los modelos de Machine Learning: el *bias* (modelos demasiado simples) y la *varianza* (también llamado overfitting, modelos demasiado complejos).
- [Preguntas frecuentes sobre Machine Learning y BigML](#).
- [Cinco consejos para empezar con Machine Learning en la empresa](#). Basado en una charla del director de ML de Uber.
- [Tutoriales de BigML](#) no escritos por BigML (bueno, algunos sí).
- [Artículos introductorios al Machine Learning](#).

# Recursos recomendados

## Vídeos

- [Vídeo tutoriales de BigML](#). Es un buen compendio de vídeos cortos. Si tienes dudas sobre algún concepto o cómo usar algo de la plataforma, no dudes en pasarte por aquí: modelos, ensembles, evaluaciones, clusters, sources, datasets... hay de todo. En inglés.

# Una consideración previa

Antes de empezar con los ejercicios quería remarcar que es muy importante entender los datos antes de crear los modelos. Los datos que vamos a usar en estos ejercicios son prácticamente autoexplicativos. Es decir, no necesitan una documentación que los explique, sino que los nombres de las variables ya dicen lo que son.

Para hacer estos ejercicios no es estrictamente necesario entenderlos, porque además de ser autoexplicativos, son datos muy preparados y que dan buenos resultados con los modelos con los que vais a trabajar. Los problemas llegan cuando los modelos no funcionan y hay que entender por qué. Y si no sabemos qué nos están contando los datos, lo tendremos muy muy complicado. Pero eso no va a suceder en este curso.

Mi recomendación es que antes de poneros con los modelos, descarguéis los datos en vuestros ordenadores y los visualicéis con Excel o con cualquier otra herramienta que hayáis estudiado ya, como Tableau. No lo he puesto explícitamente en cada ejercicio, pero **es muy deseable que lo hagáis**. Veréis también que en la plataforma BigML se generan automáticamente unos histogramas que ayudan bastante.

En el mundo real... lo siento, pero los datos están bastante “sucios” y en numerosas ocasiones no encontraréis a quien realmente sabe lo que significan. Hay que captarlos, entenderlos, limpiarlos, transformarlos y seleccionar las variables que más van a servir para el objetivo de cada proyecto. Estas fases son más importantes que las propias de Machine Learning.

¡Suerte!

# Índice

- Recursos recomendados.
- **Ejercicio 1. Ver un vídeo y contestar a las preguntas.**

# Ejercicio 1. Ver un video y contestar a las preguntas

- Video: [Cómo usar Machine Learning en el mundo real](#). Webinar propio sobre un caso de uso real. Se explica cómo se hace un proyecto de principio a fin, poniendo énfasis en la fase de análisis, limpieza y estructuración de los datos, previo a cualquier aproximación con los modelos.
- Cuestionario:
  - ¿Cuáles son las fases de un proyecto de Machine Learning?
  - ¿En qué fases se emplea mas tiempo?
  - Pon 3 ejemplos de limpieza de datos
  - ¿Qué es el Feature Engineering?
  - Indica si esta afirmación es cierta o falsa y por qué: la calidad de un modelo de Machine Learning solo se puede medir cuando ya se ha puesto en producción y se comparan las predicciones con lo que ha sucedido en la realidad.



agonzalez@cleverdata.io