

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 2. Mi primer modelo.

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 2. Mi primer modelo.**

Ejercicio 2. Mi primer modelo. Enunciado

- Objetivo: crear un primer modelo y ver su calidad.
- Vamos a crear un modelo que sirve para saber si el sentido de un tuit es positivo, negativo o neutral.
- Sube a BigML el dataset de análisis de sentimiento:

<https://cleverdata.io/csv/Sentiment.csv>
- Haz un RANDOM SPLIT 80/20.
- En el dataset resultante, pon como “no preferred” los siguientes campos: “sentiment_confidence”, “name” (deja los otros que vienen desactivados por defecto).
- Pon “sentiment” como campo objetivo y crea un modelo de árbol único ”1-CLICK MODEL” con el 80%.



Ejercicio 2. Mi primer modelo. Enunciado



- Evalúa la calidad del modelo con el 20%.
- **PREGUNTA 2.1:** ¿Qué valores de calidad ofrece el Dashboard para cada clase? Rellena los valores de la tabla.

	Negative	Neutral	Positive
Accuracy			
Precision			
Recall			
Phi Coefficient			



- **PREGUNTA 2.2:** ¿Qué sentimiento detecta mejor el modelo?

Ejercicio 2. Mi primer modelo. Enunciado

- Material de apoyo:
 - ❖ Video: [Clasificación y regresión: árboles de decisión](#)
 - ❖ Vídeo: [Predicción y evaluación](#)
 - ❖ Artículo: [Machine Learning: predicciones basadas en datos con BigML](#)
- Enlaces a entregar:
 - Dataset:
 - Dataset 80%:
 - Dataset 20%:
 - Modelo Árbol 80%:
 - Evaluación Árbol:





agonzalez@cleverdata.io