

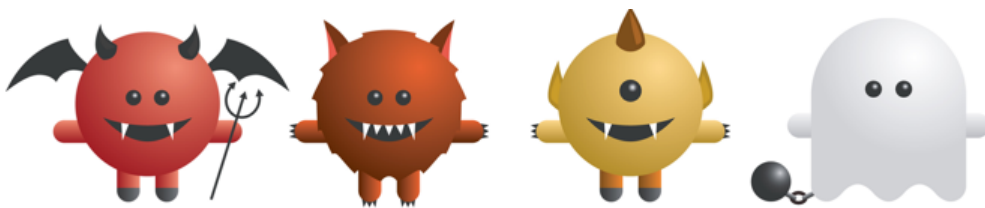
1 Dataset para las prácticas

Después de un mes haciendo observaciones científicas, y realizando mediciones cuidadosas en los pasillos de la 3ª planta del Edificio Einstein del Campus de Rabanales, bajo la supervisión de los famosos especialistas *Cazafantasmas S.A*, el Departamento de Informática ha determinado que 900 criaturas, entre demonios, fantasmas y duendes, están infestando pasillos y despachos, y asustando a nuestros alumnos científicos de datos.

Los *Cazafantasmas*, una vez agotadas las trampas disponibles y no dando fruto el almacenamiento de las criaturas, nos proponen que, dado el gremio del lugar donde estos demonios han ido a hacer de las suyas, usemos el Aprendizaje Automático como posible respuesta para desterrarlos. Así que ahora ha llegado el momento de hacer un llamamiento a los alumnos estudiantes de datos y que nos echen una mano con lo que hemos recopilado.

Hemos logrado **identificar** 371 de las 900 criaturas espantosas avistadas, pero necesitamos su ayuda para vencer a las 529 restantes, ya que sabiendo qué tipo de criaturas son podemos invocar un ritual determinado que las ahuyente, y solo un algoritmo de clasificación preciso puede lograrlo. ¡Puede usar las medidas tomadas sobre la longitud del hueso, la cantidad de carne podrida, la cantidad de falta de alma y otras características más para distinguir y extinguir a las criaturas!

¿Está preparado para el desafío?



1.1 Significado y valores de los atributos que permiten identificar a los 3 tipos de criaturas

- **id** - id of the creature
- **bone_length** - average length of bone in the creature, normalized between 0 and 1.
- **rotting_flesh** - percentage of rotting flesh in the creature between 0 and 1.
- **hair_length** - average hair length, normalized between 0 and 1.
- **has_soul** - percentage of soul in the creature between 0 and 1.
- **color** - dominant color of the creature: 'white', 'black', 'clear', 'blue', 'green', 'blood'
- **type** - target variable: 'Ghost', 'Goblin', and 'Ghoul'

2 Ficheros con los datos de las criaturas

- **dataset371.csv**: Fichero con las 371 criaturas identificadas y a partir del cual deberá entrenar a sus algoritmos para obtener modelos de generalización lo más precisos posibles, de forma que podamos aplicar esos modelos en el departamento para saber a qué tipo de criatura corresponde cada una de las 529 que no hemos sido capaces de identificar.