

Modelos Analíticos

Los modelos analíticos están compuestos de algoritmos matemáticos y estadísticos sobre los cuales nos apoyamos para poder aplicar análisis avanzados y de minería de datos. Estos algoritmos pueden ser dos tipos: Aprendizaje Supervisado y Aprendizaje no Supervisado.

Los algoritmos del tipo aprendizaje supervisado se caracteriza por contar con información que especifica cuáles datos son satisfactorios para el objetivo del aprendizaje.

En el caso de aprendizaje no supervisado, el algoritmo no cuenta con datos que definan que información es satisfactoria.

Aprendizaje Supervisado y No Supervisado

En el aprendizaje supervisado, nuestros datos están etiquetados, es decir, cada instancia de nuestro conjunto de datos tiene un atributo que es el que clasifica a esa instancia en dos o más clases. Por ejemplo, si tenemos un conjunto de datos de hipotecas, podríamos querer tener un modelo del impago de las mismas. Entonces tendríamos en nuestro conjunto de datos:

- Atributos de entrada: podrían ser en el ejemplo el principal de la hipoteca, el tipo de interés, la edad del cliente, la profesión, etc.
- Un atributo que es la "salida", etiqueta u objetivo, que en este caso sería el campo que indicase si ese cliente pagó o no la hipoteca.

Así, los algoritmos "aprenden" tratando de buscar diferencias entre las características de las instancias de clientes que pagaron y los que no. Lo fundamental de estas técnicas es que los datos para entrenar el algoritmo tienen que venir previamente clasificados.

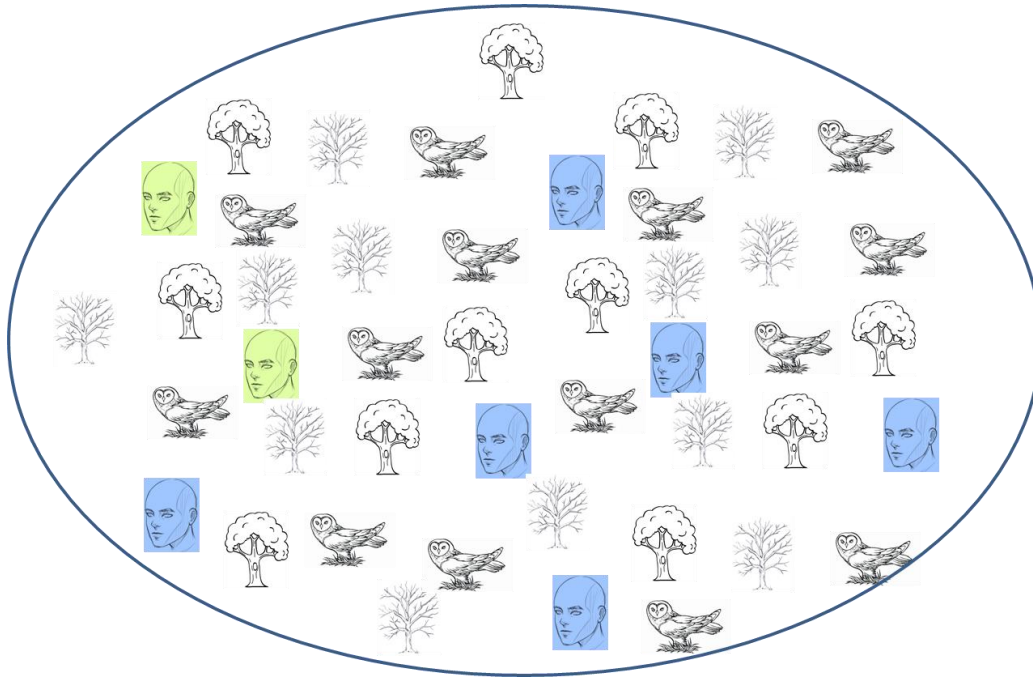
Poniendo un ejemplo visual, analicemos el siguiente conjunto de datos:



Si nos fijamos, tenemos 4 tipos de imágenes distintas. Supongamos que necesitamos identificar todos los rostros del conjunto de datos. Con un algoritmo de aprendizaje supervisado podemos lograr este objetivo. Primero identificamos manualmente una muestra de los rostros existentes, tal como se muestra a continuación con los rostros resaltados en verde:



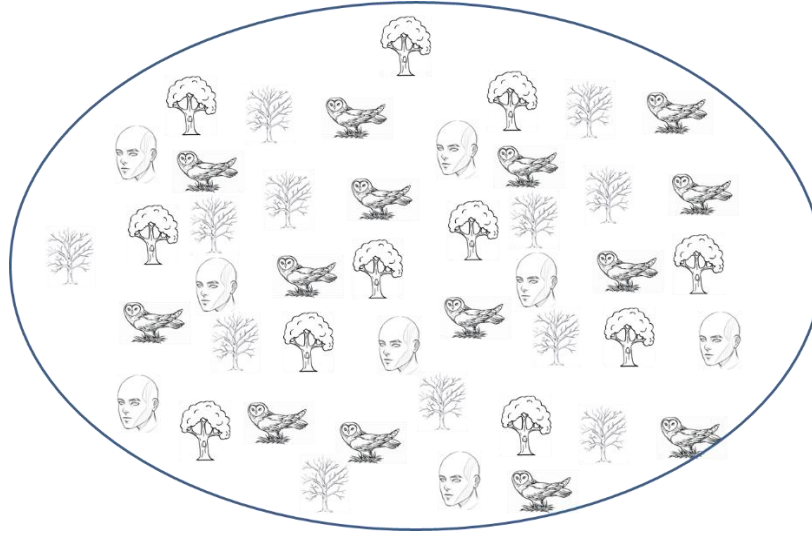
A partir de esta clasificación inicial de la data, el algoritmo es capaz de identificar los rostros restantes en el conjunto de datos (los rostros resaltados en azul a continuación).



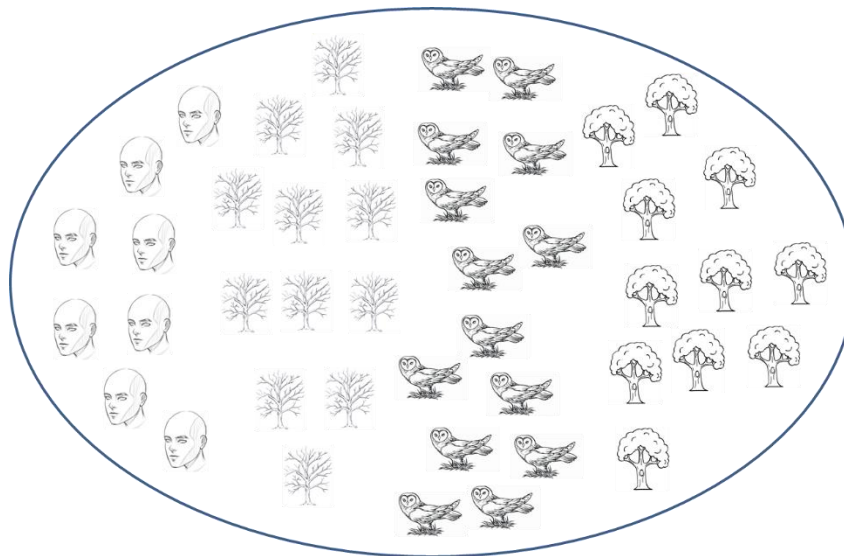
En el caso del aprendizaje no supervisado, no se tiene (o no se considera) esa etiqueta que clasifica a priori a los datos. Por ejemplo, podríamos querer simplemente buscar segmentos de clientes, es decir, buscar grupos de clientes homogéneos para crear productos personalizados, pero no sabemos a priori cuántos grupos habrá ni sus características. Este es un ejemplo de "agrupamiento" (clustering).

Es importante entender que el uso de aprendizaje supervisado o no supervisado depende de los objetivos de nuestro estudio, y no solo del conjunto de datos. Por ejemplo, con el conjunto de datos impago de hipotecas antes mencionado, podría también utilizarse para un propósito de agrupamiento, excluyendo por ejemplo los clientes que no pagaron de los datos. Aunque los datos utilizados vienen del mismo conjunto de datos, el propósito es completamente diferente.

Si retomamos el ejemplo visual utilizado anteriormente, veamos el conjunto de datos nuevamente:



Un algoritmo de aprendizaje no supervisado es incapaz de decirnos si una imagen dada es un rostro o no pero sí podría, por ejemplo, clasificar las imágenes entre aquellas que contienen rostros humanos, de animales, o las que no contienen. A continuación el resultado de un algoritmo de este tipo:



La información obtenida por un algoritmo de aprendizaje no supervisado debe ser posteriormente interpretada por una persona para darle utilidad.