

Actividad 4

Filtrar mensajes de spam de teléfono móvil con el clasificador Naïve Bayes

Las empresas anunciantes utilizan Short Message Service (SMS) para dirigirse a los consumidores potenciales con publicidad conocida como spam SMS. Este tipo de spam es particularmente problemático porque, a diferencia del spam de correo electrónico, muchos usuarios de teléfonos móviles pagan una tarifa por cada SMS recibido. El desarrollo de un algoritmo de clasificación que pudiera filtrar el spam SMS proporcionaría una herramienta útil para los proveedores de telefonía móvil.

Usar el dataset 'sms_spam.csv', que se corresponde con datos adaptados de la Colección de Spam SMS en <http://www.dt.fee.unicamp.br/~tiago/smsspamcollection/>

Este conjunto de datos incluye el texto de los mensajes SMS junto con una etiqueta que indica si el mensaje es deseado o no. Los mensajes basura se etiquetan como '*spam*', mientras que los mensajes legítimos se etiquetan como '*ham*'.

El clasificador calculará la probabilidad de '*spam*' y '*ham*' dada la evidencia proporcionada por todas las palabras que aparezcan en un mensaje.

Crear la nube de palabras correspondiente a cada una de las clases, de manera que se evidencien las palabras clave fuertes para diferenciar ambas clases.