

PRÁCTICA 3: MINERÍA DE TEXTO PARA DETECTAR TEMÁTICAS DONDE SURGEN MÁS DUDAS Y SENTIMIENTOS DE LOS ALUMNOS

Como se ha visto a lo largo de las clases teóricas, se puede realizar la detección de diferentes indicadores de los alumnos que sirven a modo de modelo de usuario, a través de muy diversas técnicas o combinaciones de ellas, como por ejemplo minería de procesos, métodos probabilísticos, minería de textos o machine learning. En la práctica 1 hubo la oportunidad de aplicar métodos basados en minería de proceso sencilla para la detección de trampas. En la práctica 2 hubo la oportunidad de aplicar métodos probabilísticos para la estimación de la habilidad, mediante la Teoría de Respuesta al Ítem o las Redes Bayesianas. Además, en una práctica posterior, se aplicarán métodos de “machine learning” para propósitos predictivos. Como uno de los fines es la utilización de diferentes métodos para la aplicación de Inteligencia Artificial en Educación, en esta práctica nos centraremos en la aplicación de técnicas de minería de datos para la detección de ciertos indicadores que se pueden extraer de los alumnos:

TAREA 1: A través del siguiente enlace se encuentra disponible un corpus de datos de mensajes de foros de la comunidad Kaggle, un corpus que es público (licencia CCO de dominio público) y se puede observar en <https://www.kaggle.com/datasets/yashvi/kaggle-feedback-comments> . En primer lugar, nos debemos familiarizar con este corpus de datos y cómo son sus mensajes.

Se encuentran dos ficheros: uno con las temáticas de los foros, y otro con los propios posts en los foros (tanto los que crean un hilo como sus respuestas).

Aunque el propósito de este foro no es exclusivamente educativo, se puede ver como un medio de aprendizaje en una comunidad abierta. Sobre estos mensajes en los foros nos centraremos en esta práctica.

Este corpus de datos se debe crear en R e importar los datos.

TAREA 2: La minería de textos son un conjunto de técnicas para tomando texto poder extraer una serie de conclusiones o incluso ser capaz de interpretarlo y proporcionar un feedback o diálogo con los usuarios.

La minería de texto se puede utilizar con fines educativos, como por ejemplo para saber qué temáticas tienen más dudas los alumnos, cuál es el nivel de los alumnos en diferentes áreas, detectar habilidades meta-cognitivas, o detectar los sentimientos de los alumnos.

En esta tarea se debe determinar cuáles son las temáticas más habituales que se hablan en este foro (se debe realizar de manera independiente al fichero de temáticas). Es decir, se deben extraer las temáticas sin saber inicialmente cuáles son los temas de que se tratan. También se debe determinar cuáles son las temáticas en las que hay más dudas.

TAREA 3: En esta tarea se debe determinar el sentimiento de los usuarios para cada uno de los posts del foro (positivo, negativo o neutro). Así mismo, se deben ver las temáticas donde se han producido más sentimientos negativos.

TAREA 4: En esta tarea, aprovechando la información que se proporciona en el fichero de datos, se deben realizar otras aplicaciones con los datos utilizando minería de texto.

Para la realización de esta práctica hay libertad en las técnicas y métodos a utilizar de minería de texto. A modo de ejemplo, en el siguiente enlace se pueden encontrar cuatro cursos sobre el uso de R para minería de textos aplicando “bag of words”:

<https://app.datacamp.com/learn/courses/text-mining-with-bag-of-words-in-r>