

## Taller

Esta práctica tiene como objetivo analizar el sentimiento de dos reseñas de película una vez entrenado un modelo con una base de datos etiquetada usando aprendizaje automático.

La idea es poder clasificar las reseñas de películas en un sentimiento positivo frente al negativo. El modelo debe ser entrenado con datos que sean: a) relacionados con el tema, es decir, reseñas de películas; y b) revele información sobre el sentimiento de cada revisión, es decir, positiva frente a negativa.

Un conjunto de datos que cumple con estas expectativas se puede identificar en Polarity Dataset v2.0 es Movie Review Data disponible en:

<http://www.cs.cornell.edu/people/pabo/movie-review-data/>

Este conjunto de datos fue generado dentro de un proyecto de investigación en la Universidad de Cornell. Este repositorio contiene 1000 revisiones positivas y 1000 críticas negativas.

Después de descargar los datos, procese cada uno de ellos para extraer los términos relacionados utilizando la opción de creación de vectores TF-IDF. Se recomienda ejecutar un proceso previo de pre-procesamiento donde al menos aplique tokenización, stem y eliminación de stop-words. Divida el conjunto de datos en 70% para entrenamiento y use los mismos para entrenar un clasificador SVM dentro del operador de validación (validation). Almacene tanto el modelo entrenado como los términos procesados en un repositorio

Finalmente, se pide aplicar el modelo entrenado a dos revisiones de películas que no formaban parte de los datos de entrenamiento. Para la aplicación del modelo use la siguiente información

Los términos utilizados en el modelo entrenado

Los términos de la fuente original que fueron pre-procesados y luego agregue las dos revisiones nuevas a las cuales se debe aplicar las mismas transformaciones (preprocesamiento) que en el caso de los datos entrenados anteriormente.

Finalmente, se aplica el modelo entrenado a la matriz de documento de término resultante y se obtiene un puntaje sentimiento para cada revisión de las películas ejemplo