

Objetivos

1. Estudio del estado del arte en el campo de la minería de opinión basada en Big Data en plataformas de microblogging.
2. Desarrollo y aplicación de una metodología de preprocesado de datos eficaz para conjuntos de textos provenientes de plataformas de microblogging.
3. Aplicación de técnicas de minería de datos descriptiva para obtener patrones interesantes en los datos.
4. Obtención y salvado de un corpus de datos de gran tamaño que permitan elaborar la experimentación y validación del sistema final.
5. Experimentación, análisis de resultados y comparación de los mismos con posibles eventos políticos y sociales.
6. Puesta en valor del sistema mediante técnicas de visualización dinámicas.

Planificación del proyecto



Timing: Se dividió el proyecto en obtención de información, estudio estado del arte, obtención del data set, preprocesado, minería de datos y exploración de resultados, **outline** clásico de ciencia de datos.



Personal: El autor del proyecto y las directoras del mismo.



Hardware: Mac Pro 2,6 GHz Intel Core i5 8 GB 1600 MHz DDR3. Cluster con 4 nodos de Intel Xeon E5-2665 de 8 núcleos.



Software: TexShop, TextEdit, Rstudio, Spark, PySpark, PyCharm, nltk, Numbers, MongoDB.