

# SIW PL11-12. Datamining.

Hugo Fonseca Díaz  
uo258318@uniovi.es

*Escuela de Ingeniería Informática. Universidad de Oviedo.*

10 de diciembre de 2020

## 1. Objetivo

El objetivo de esta práctica es crear un clasificador que realice predicciones de si el índice bursátil *Dow Jones* va a subir o a bajar según una serie de noticias internacionales registradas ese día. Para crear dicho clasificador usaremos la herramienta *Orange* [1].

## 2. Diseño

Para dicho clasificador se ha creado un corpus formado por los datos extraídos de una hoja de cálculo. Al preprocesarse dicho corpus se le añade una lista de palabras vacías incluidas en el fichero *stopwords.txt* creado por nosotros. Después se crea una bolsa de palabras con una frecuencia de términos y un cálculo del IDF. Dicha bolsa se conecta con varios modelos que a su vez muestran los resultados de sus cálculos en un nanograma. Este nanograma nos muestra las palabras más importantes a la hora de clasificar si una noticia hará que el índice *Dow Jones* baje o suba.

## 3. Resultados

Una vez observados los nanogramas obtenidos, probamos y clasificamos los diferentes modelos. La regresión logística nos da un porcentaje de acierto por debajo del 30 % (sobre el 28 %). El bayesiano ingenuo nos da un porcentaje de acierto de casi el 50 % (sobre un 49 %).

## Referencias

[1] Página de *Orange*, <https://orange.biolab.si/>. Última vez accedido 10 de diciembre de 2020.