



Universidad  
de La Laguna

Escuela Técnica Superior  
de Ingeniería Informática

Inteligencia Artificial Avanzada

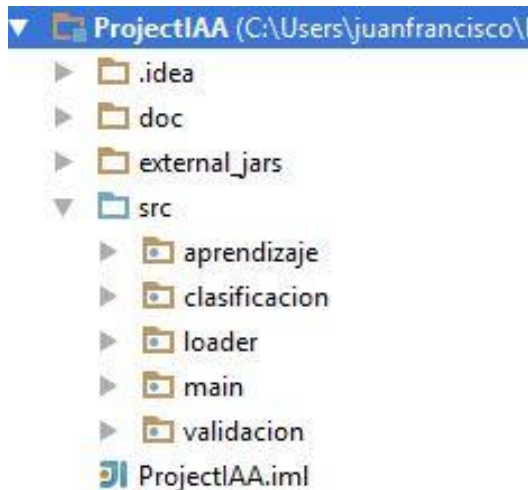
# Proyecto Final

Análisis de Sentimiento en Lenguaje Natural

Juan F. González Ramos

Gonzalo García Martín

## ESTRUCTURA DEL PROYECTO



La estructura del proyecto corresponde a lo siguiente:

**.idea:** corresponde a la configuración del proyecto por parte del IDE **IntelliJ IDEA**

**doc:** contiene la documentación Javadoc del proyecto

**external\_jars:** contiene las librerías que por defecto no están incluidas en el JDK, en este caso es la librería commons para la manipulación de ficheros.

**src:** esta carpeta contiene el código fuente del proyecto y tiene la siguiente estructura.

- Aprendizaje
  - Por cada palabra de vocabulario.txt, su frecuencia en el corpus y una estimación del logaritmo de su probabilidad mediante suavizado laplaciano con tratamiento de palabras desconocidas. Las palabras en los ficheros de aprendizaje estarán ordenadas alfabéticamente.
- Clasificación
  - En esta parte se clasificarán los documentos presentes en un corpus.
  - Toma como entrada las estimaciones de probabilidad de cada palabra y un corpus con documentos a clasificar y devuelve los documentos clasificados en un fichero clasificación.txt
- Loader
  - Crea un corpus con nombre corpus<pos o neg>.txt que una todos los documentos con opinión positiva y otro corpus con todos los documentos con opinión negativa.
  - Halla el vocabulario del problema. Para ello examina el fichero corpustodo.txt y obtén las palabras del vocabulario a partir del texto.
- Main
  - Contiene la ventana de bienvenida que observa el usuario al comenzar la ejecución del programa
- Validación
  - En esta parte se estimará el rendimiento del sistema.
  - Toma el fichero clasificación.txt y otro fichero con el mismo formato llamado clasificación\_ok.txt y escribir el porcentaje de errores cometidos en la clasificación (con qué porcentaje la clase para clasificación.txt no coincide con clasificación\_ok.txt) en el fichero error.txt

## **REPARTO DE TAREAS Y COLABORACION ENTRE ALUMNOS.**

La generación del código fuente para este proyecto se hizo de forma colaborativa mediante el pair programming durante varios días seguidos permitiendo a cada alumno asumir no solo la misma carga de trabajo sino que la estructura de clases fue diseñada por ambos alumnos.

Las tareas secundarias como la documentación de código usando Javadoc o el control de versiones usando GitHub se repartieron entre ambos alumnos de forma que la carga de trabajo fuese equitativa.