



**Universidad  
Europea**

LAUREATE INTERNATIONAL UNIVERSITIES

**UNIVERSIDAD EUROPEA DE MADRID**

**ESCUELA DE ARQUITECTURA, INGENIERÍA Y DISEÑO**

**GRADO EN INGENIERÍA INFORMÁTICA**

**PROYECTO DE COMPUTACIÓN I**

# **DISEÑO DEL PLAN DEL PROYECTO**

**DIEGO ABDUL MASSIH LOPEZ**

**MIGUEL ROMERO GARCÍA**

**JOSÉ ACEITUNO GARCÍA**

**DIEGO GIL EDO**

**CURSO 2018-2019**

# ÍNDICE

|   |           |
|---|-----------|
| <b>INTRODUCCIÓN</b>                                 | <b>2</b>  |
| <b>ORGANIZACIÓN DEL EQUIPO</b>                      | <b>2</b>  |
| 2.1 ACTORES DEL SISTEMA                             | 3         |
| <b>OBJETIVOS DEL SISTEMA</b>                        | <b>3</b>  |
| <b>ESPECIFICACIÓN DE REQUISITOS Y CASOS DE USO</b>  | <b>4</b>  |
| REQUISITOS FUNCIONALES                              | 4         |
| REQUISITOS NO FUNCIONALES                           | 7         |
| CASOS DE USO  | 8         |
| <b>ESPECIFICACIÓN DE REQUISITOS SOFTWARE.</b>       | <b>14</b> |
| DATOS DE ENTRADA Y FORMATOS                         | 14        |
| TRANSFORMACIONES A LLEVAR A CABO CON LOS DATOS      | 14        |
| DESCRIPCIÓN DEL ALGORITMO A USAR                    | 15        |
| DESCRIPCIÓN DEL “CROSS VALIDATION”                  | 15        |
| DESCRIPCIÓN DE LA INTERFAZ DE USUARIO               | 16        |
| DESCRIPCIÓN DE LA SALIDA QUE GENERARÁ LA APLICACIÓN | 18        |
| DESCRIPCIÓN DEL ENTREGABLE QUE RECIBIRÁ EL CLIENTE  | 18        |
| <b>DISEÑO TÉCNICO PRELIMINAR.</b>                   | <b>19</b> |
| DIAGRAMA DE COMPONENTES DEL SISTEMA                 | 19        |
| DESCRIPCIÓN DE LAS LIBRERÍAS                        | 19        |
| Panda   | 19        |
| Sickit-learn  | 19        |
| Matplotlib  | 19        |
| Os  | 19        |
| Numpy   | 20        |
| PyQT 5  | 20        |
| LangDetect  | 20        |
| Pycountry   | 20        |
| DIAGRAMA DE CLASES                                  | 20        |
| <b>PLAN DE PROYECTO</b>                             | <b>21</b> |
| DIAGRAMA EDT  | 21        |
| DIAGRAMA DE GANTT                                   | 21        |
| PRESUPUESTO   | 22        |
| TABLA DE DESCRIPCIÓN DE FUNCIONALIDADES             | 22        |
| <b>MANUAL DE INSTALACIÓN</b>                        | <b>22</b> |
| <b>TABLA DE PRUEBAS</b>                             | <b>23</b> |
| <b>BIBLIOGRAFÍA</b>                                 | <b>25</b> |

## 1. INTRODUCCIÓN

La finalidad de este trabajo es la de colaborar con un cliente, para ofrecerle una solución en forma de software, que sea capaz de clasificar las opiniones/reseñas que dejan sus clientes en internet sobre su negocio, y de esta forma saber cuáles son sus puntos fuertes y cuáles son sus puntos débiles, para ser capaz de gestionarlo mejor.

Actualmente, internet, se ha convertido en una tecnología utilizada e imprescindible por millones de personas al día. Antes, el éxito o el fracaso de un negocio, estaba asociado al boca a boca de las personas, sin embargo, actualmente y en el mundo del futuro, esto ha cambiado muchísimo, y es que, hoy en día, la mayoría de valoraciones se mueven vía online. Es por ello, que estas opiniones en internet tienen mucho peso en el rumbo y el futuro de un negocio, si estas opiniones son buenas, es más probable que la clientela acuda al negocio en búsqueda de información. En cambio, si las opiniones son malas, es menos probable que el negocio tenga éxito.

Por ello, el cliente requiere una aplicación la cual le permita clasificar las opiniones de sus clientes sobre sus servicios en buenas, neutras y malas, aunque está especialmente interesado en las buenas y malas, puesto que de esta manera podrá mejorar el aspecto de su negocio de cara al público.

## 2. ORGANIZACIÓN DEL EQUIPO

El equipo estará formado por cuatro estudiantes de Ingeniería Informática e Ingeniería Informática y DYCRE (Dirección y Creación de Empresas). Estos son:

- José Aceituno (*José A.*), estudiante de tercero de Ingeniería Informática y DYCRE.
- Diego Gil Edo (*Diego G.*), estudiante de tercero de Ingeniería Informática y DYCRE.
- Diego Abdul (*Diego A.*), estudiante de tercero de Ingeniería Informática.
- Miguel Romero (*Miguel R.*), estudiante de cuarto de Ingeniería Informática.

## 2.1 ACTORES DEL SISTEMA

En este apartado, se incluyen los distintos roles que se incluirán dentro del sistema.

### **ACT- 01** Usuario General

|                    |   |
|--------------------|---|
| <i>Descripción</i> | <i>Este actor es el único actor del gestor.</i> |
|--------------------|---|

## 3. OBJETIVOS DEL SISTEMA

### **OBJ-01** Elaboración de un sistema de clasificación de opiniones

|                    |  |
|--------------------|--|
| <i>Descripción</i> | El sistema deberá clasificar las opiniones sobre el negocio de un cliente para saber si estas son buenas, malas o neutras. |
| <i>Estabilidad</i> | Alta   |
| <i>Comentarios</i> | Ninguno  |

### **OBJ-02** Visualización de estadísticas sobre la clasificación de textos

|                    |   |
|--------------------|---|
| <i>Descripción</i> | El sistema deberá mostrar las estadísticas generadas por la clasificación de textos |
| <i>Estabilidad</i> | Alta  |
| <i>Comentarios</i> | Ninguno   |

### **OBJ-03** Elaboración proceso de entrenamiento para el algoritmo elegido

|                    |   |
|--------------------|---|
| <i>Descripción</i> | El sistema permitirá crear nuevos entrenamientos en base a los textos de entrada y las opciones elegidas. |
| <i>Estabilidad</i> | Alta  |
| <i>Comentarios</i> | Ninguno   |

**OBJ-04** Elaboración proceso de clasificación a partir del entrenamiento

|             |  |
|-------------|--|
| Descripción | El sistema permitirá generar nuevas clasificaciones a partir de entrenamientos creados previamente |
| Estabilidad | Alta   |
| Comentarios | Ninguno  |

## 4. ESPECIFICACIÓN DE REQUISITOS Y CASOS DE USO

### 4.1 REQUISITOS FUNCIONALES

**RF-01** Clasificación de valoraciones

|                      |  |
|----------------------|--|
| Objetivos Asociados  | OBJ-01, OBJ-02, OBJ-03, OBJ-04   |
| Requisitos Asociados |  |
| Descripción          | El sistema permitirá, a través de un algoritmo entrenado previamente, la clasificación de valoraciones según sean buenas, neutras o malas. |
| Datos específicos    |  |

**RF-02** Selección de directorios y textos

|                      |   |
|----------------------|---|
| Objetivos Asociados  | OBJ-03, OBJ-04  |
| Requisitos Asociados |   |
| Descripción          | El sistema permitirá que usuario seleccione el directorio donde se encuentren los textos que desee.           |
| Datos específicos    | <ul style="list-style-type: none"><li>• Nombre del directorio</li><li>• Tipo de archivos soportados</li></ul> |

### **RF-03** Selección del algoritmo

|                      |  |
|----------------------|--|
| Objetivos Asociados  | OBJ-03   |
| Requisitos Asociados |  |
| Descripción          | <i>El sistema permitirá la selección de un algoritmo deseado dentro de una lista de algoritmos, para la fase de entrenamiento.</i> |
| Datos específicos    | <ul style="list-style-type: none"><li>• Listado de algoritmos</li></ul>  |

### **RF-04** Entrenamiento del algoritmo

|                      |  |
|----------------------|--|
| Objetivos Asociados  | OBJ-01, OBJ-03, OBJ-04   |
| Requisitos Asociados |  |
| Descripción          | <i>El sistema permitirá el entrenamiento del algoritmo de elegido, para posteriormente poder clasificar las valoraciones.</i>      |
| Datos específicos    | <ul style="list-style-type: none"><li>• Idioma de las valoraciones</li><li>• Lista de paradas de palabras ('Stop-Words')</li></ul> |

### **RF-05** Visualización e interpretación de resultados

|                      |   |
|----------------------|---|
| Objetivos Asociados  | OBJ-02, OBJ-04  |
| Requisitos Asociados |   |
| Descripción          | <i>El sistema permitirá la visualización de e interpretación de los resultados de las clasificaciones, adicionalmente podrá ser mostrado mediante gráficas.</i> |
| Datos específicos    | <ul style="list-style-type: none"><li>• Resultados de las clasificaciones</li></ul>   |

## **RF-06** Selección de idioma

|                             |   |
|-----------------------------|---|
| <i>Objetivos Asociados</i>  | OBJ-02, OBJ-03, OBJ-04  |
| <i>Requisitos Asociados</i> |   |
| <i>Descripción</i>          | <i>El sistema permitirá al usuario seleccionar el idioma con el que desea entrenar a los algoritmos, pudiendo elegir entre castellano e inglés.</i> |
| <i>Datos específicos</i>    | <ul style="list-style-type: none"><li>● <i>Listado de idiomas</i></li></ul>   |

## 4.2 REQUISITOS NO FUNCIONALES

### **NFR-01** Multiplataforma

|                            |  |
|----------------------------|--|
| <i>Objetivos Asociados</i> | -  |
| <i>Descripción</i>         | <i>El sistema deberá ser capaz de correr tanto en Windows como en Linux.</i> |
| <i>Comentarios</i>         | ninguno  |

### **NFR-02** Usabilidad

|                            |  |
|----------------------------|--|
| <i>Objetivos Asociados</i> | -  |
| <i>Descripción</i>         | <i>El sistema debe ser sencillo y atractivo para el usuario.</i> |
| <i>Comentarios</i>         | ninguno  |

### **NFR-03** Eficiencia

|                            |   |
|----------------------------|---|
| <i>Objetivos Asociados</i> | -   |
| <i>Descripción</i>         | <i>El sistema debe estar lo más optimizado posible a través de algoritmos eficientes.</i> |
| <i>Comentarios</i>         | ninguno   |



### 4.3 CASOS DE USO

#### CU-01 Configuración de datos (salida)

| Objetivos Asociados  | OBJ-01, OBJ-02   |  |      |        |   |  |   |  |
|----------------------|--|--|------|--------|---|--|---|--|
| Requisitos asociados | RF01, RF05   |  |      |        |   |  |   |  |
| Descripción          | El usuario deberá especificar en qué parte de su disco duro almacenará la salida que ofrece nuestra aplicación.  |  |      |        |   |  |   |  |
| Precondición         | El algoritmo debe de haber sido entrenado correctamente con las valoraciones ya clasificadas.  |  |      |        |   |  |   |  |
| Secuencia normal     | <table><tr><th>Paso</th><th>Acción</th></tr><tr><td>1</td><td>El usuario busca en su equipo la carpeta en la que desee guardar las valoraciones clasificadas por el algoritmo.</td></tr><tr><td>2</td><td>El usuario indica la carpeta de valoraciones clasificadas por el algoritmo al sistema.</td></tr></table> |  | Paso | Acción | 1 | El usuario busca en su equipo la carpeta en la que desee guardar las valoraciones clasificadas por el algoritmo. | 2 | El usuario indica la carpeta de valoraciones clasificadas por el algoritmo al sistema. |
| Paso                 | Acción   |  |      |        |   |  |   |  |
| 1                    | El usuario busca en su equipo la carpeta en la que desee guardar las valoraciones clasificadas por el algoritmo.   |  |      |        |   |  |   |  |
| 2                    | El usuario indica la carpeta de valoraciones clasificadas por el algoritmo al sistema.   |  |      |        |   |  |   |  |
| Postcondición        | La carpeta con las valoraciones es correctamente indicada para su posterior entrenamiento.   |  |      |        |   |  |   |  |
| Excepciones          | -  |  |      |        |   |  |   |  |
| Estabilidad          | Alta   |  |      |        |   |  |   |  |
| Comentarios          | Ninguno  |  |      |        |   |  |   |  |

## CU-02

## Configuración de datos (entrada)

| Objetivos Asociados  | OBJ-01, OBJ-02  |  |      |        |   |  |   |   |
|----------------------|---|--|------|--------|---|--|---|---|
| Requisitos asociados | RF02, RF-03, RF-06  |  |      |        |   |  |   |   |
| Descripción          | El usuario deberá especificar de qué parte de su disco duro la aplicación deberá de coger los textos para proceder con la clasificación.  |  |      |        |   |  |   |   |
| Precondición         | Existencia de una carpeta con valoraciones definidas (buenas, malas y neutras).   |  |      |        |   |  |   |   |
| Secuencia normal     | <table><tr><th>Paso</th><th>Acción</th></tr><tr><td>1</td><td>El usuario busca en su equipo la carpeta con las valoraciones clasificadas</td></tr><tr><td>2</td><td>El usuario indica la carpeta de valoraciones al sistema</td></tr></table> |  | Paso | Acción | 1 | El usuario busca en su equipo la carpeta con las valoraciones clasificadas | 2 | El usuario indica la carpeta de valoraciones al sistema |
| Paso                 | Acción  |  |      |        |   |  |   |   |
| 1                    | El usuario busca en su equipo la carpeta con las valoraciones clasificadas  |  |      |        |   |  |   |   |
| 2                    | El usuario indica la carpeta de valoraciones al sistema   |  |      |        |   |  |   |   |
| Postcondición        | La carpeta con las valoraciones es correctamente indicada para su posterior entrenamiento.  |  |      |        |   |  |   |   |
| Excepciones          | -   |  |      |        |   |  |   |   |
| Estabilidad          | Alta  |  |      |        |   |  |   |   |
| Comentarios          | Ninguno.  |  |      |        |   |  |   |   |

## CU-03

## Visualización de resultados

| Objetivos Asociados  | OBJ-01, OBJ-02  |  |      |        |   |   |
|----------------------|---|--|------|--------|---|---|
| Requisitos asociados | RF01, RF05, RF06  |  |      |        |   |   |
| Descripción          | El usuario tendrá la opción de visualizar los resultados (la salida) de nuestro clasificador.   |  |      |        |   |   |
| Precondición         | Las valoraciones deben de haber sido clasificadas por el algoritmo seleccionado.  |  |      |        |   |   |
| Secuencia normal     | <table><tr><th>Paso</th><th>Acción</th></tr><tr><td>1</td><td>El usuario selecciona el algoritmo para clasificar las valoraciones</td></tr></table> |  | Paso | Acción | 1 | El usuario selecciona el algoritmo para clasificar las valoraciones |
| Paso                 | Acción  |  |      |        |   |   |
| 1                    | El usuario selecciona el algoritmo para clasificar las valoraciones   |  |      |        |   |   |
| Postcondición        | El sistema muestra visualmente los resultados correspondientes a las valoraciones clasificadas.   |  |      |        |   |   |
| Excepciones          | -   |  |      |        |   |   |
| Estabilidad          | Alta.   |  |      |        |   |   |
| Comentarios          | Ninguno.  |  |      |        |   |   |

## CU-04 Selección de algoritmos

| Objetivos Asociados  | OBJ-03, OBJ-04   |  |      |        |   |   |   |   |
|----------------------|--|--|------|--------|---|---|---|---|
| Requisitos asociados | RF-03, RF-04   |  |      |        |   |   |   |   |
| Descripción          | El usuario podrá elegir el algoritmo con el que desea ejecutar la clasificación de las opiniones.  |  |      |        |   |   |   |   |
| Precondición         | Existencia de unas opiniones clasificadas para que el algoritmo pueda entrenar.  |  |      |        |   |   |   |   |
| Secuencia normal     | <table><tr><th>Paso</th><th>Acción</th></tr><tr><td>1</td><td>El programa muestra al usuario las opciones que puede elegir para la fase de entrenamiento.</td></tr><tr><td>2</td><td>El usuario indica el algoritmo que desee.</td></tr></table> |  | Paso | Acción | 1 | El programa muestra al usuario las opciones que puede elegir para la fase de entrenamiento. | 2 | El usuario indica el algoritmo que desee. |
| Paso                 | Acción   |  |      |        |   |   |   |   |
| 1                    | El programa muestra al usuario las opciones que puede elegir para la fase de entrenamiento.  |  |      |        |   |   |   |   |
| 2                    | El usuario indica el algoritmo que desee.  |  |      |        |   |   |   |   |
| Postcondición        | Las valoraciones están correctamente asignadas.  |  |      |        |   |   |   |   |
| Excepciones          | -  |  |      |        |   |   |   |   |
| Estabilidad          | Alta.  |  |      |        |   |   |   |   |
| Comentarios          | Ninguno.   |  |      |        |   |   |   |   |

## CU-05

## Selección de listas de parada (Stop-Words)/Idioma

| Objetivos Asociados  | OBJ-01, OBJ-02  |   |      |        |   |   |   |  |
|----------------------|---|---|------|--------|---|---|---|--|
| Requisitos asociados | NFR-02, NFR-03  |   |      |        |   |   |   |  |
| Descripción          | El usuario podrá elegir la lista de paradas con la que desee ejecutar la clasificación de las opiniones, cambiándose automáticamente cuando elija otro idioma de la lista.  |   |      |        |   |   |   |  |
| Precondición         | Existencia de unas opiniones ya clasificadas para que el programa pueda entrenar.   |   |      |        |   |   |   |  |
| Secuencia normal     | <table><tr><th>Paso</th><th>Acción</th></tr><tr><td>1</td><td>El programa muestra al usuario las opciones que puede elegir para la fase de entrenamiento.</td></tr><tr><td>2</td><td>El usuario indica la lista de parada que desee o bien introducir una personal.</td></tr></table> |   | Paso | Acción | 1 | El programa muestra al usuario las opciones que puede elegir para la fase de entrenamiento. | 2 | El usuario indica la lista de parada que desee o bien introducir una personal. |
|                      | Paso  | Acción  |      |        |   |   |   |  |
|                      | 1   | El programa muestra al usuario las opciones que puede elegir para la fase de entrenamiento. |      |        |   |   |   |  |
|                      | 2   | El usuario indica la lista de parada que desee o bien introducir una personal.              |      |        |   |   |   |  |
| Postcondición        | Las valoraciones están correctamente asignadas.   |   |      |        |   |   |   |  |
| Excepciones          | -   |   |      |        |   |   |   |  |
| Estabilidad          | Alta.   |   |      |        |   |   |   |  |
| Comentarios          | Ninguno.  |   |      |        |   |   |   |  |

## CU-06

## Visualización avanzada de resultados

| <i>Objetivos Asociados</i>  | OBJ-01, OBJ-02  |      |        |   |   |   |   |
|-----------------------------|---|------|--------|---|---|---|---|
| <i>Requisitos asociados</i> | RF-01, RF-05  |      |        |   |   |   |   |
| <i>Descripción</i>          | El usuario podrá tener una vista de la precisión del clasificador por medio de porcentajes y gráficos de línea o barras.  |      |        |   |   |   |   |
| <i>Precondición</i>         | Existencia de unos datos clasificados por el programa.  |      |        |   |   |   |   |
| <i>Secuencia normal</i>     | <table> <tr> <th>Paso</th><th>Acción</th></tr> <tr> <td>1</td><td>El programa muestra al usuario la clasificación de todas opiniones.</td></tr> <tr> <td>2</td><td>El usuario indica la visualización deseada.</td></tr> </table> | Paso | Acción | 1 | El programa muestra al usuario la clasificación de todas opiniones. | 2 | El usuario indica la visualización deseada. |
| Paso                        | Acción  |      |        |   |   |   |   |
| 1                           | El programa muestra al usuario la clasificación de todas opiniones.   |      |        |   |   |   |   |
| 2                           | El usuario indica la visualización deseada.   |      |        |   |   |   |   |
| <i>Postcondición</i>        | El usuario sabe la precisión exacta de la clasificación.  |      |        |   |   |   |   |
| <i>Excepciones</i>          | -   |      |        |   |   |   |   |
| <i>Estabilidad</i>          | Alta.   |      |        |   |   |   |   |
| <i>Comentarios</i>          | Ninguno.  |      |        |   |   |   |   |

## 5. ESPECIFICACIÓN DE REQUISITOS SOFTWARE.

### 5.1 DATOS DE ENTRADA Y FORMATOS

Los datos de entrada son archivos de texto(.txt) que contienen las valoraciones que el usuario quiera clasificar. El formato de estos archivos será el siguiente:

- Primera línea del archivo: Valoración numérica del 1 al 10 con solo un decimal y usando un punto para separar los decimales.
- Segunda línea y siguientes: Texto de la valoración.

El programa podrá seleccionar las carpetas donde se encuentren estas valoraciones a través de un explorador de directorios.

Adicionalmente, y de cara al futuro, se considera un mecanismo de descompresión de archivos .zip integrado en el programa, para poder elegir uno o varias carpetas de un solo intento. Evidentemente, zip tendría que estar compuesto únicamente por carpetas y estas carpetas por archivos de texto con el formato mencionado anteriormente. Los nombres de estas carpetas tendrían que ser 'Buenas', 'Neutras' o 'Malas' para que el programa las pudiese clasificar automáticamente, si no se diese ese supuesto, el programa preguntará al usuario la naturaleza de cada carpeta.

Otra mejora propuesta contemplada de cara al futuro, sería la de guardar cada uno de los datos introducidos en una base de datos, con el objetivo de tener un dataset global cada vez más grande y ordenado, de donde los algoritmos puedan entrenar de una manera mejor y eficiente, con el objetivo de dar resultados más precisos.

### 5.2 TRANSFORMACIONES A LLEVAR A CABO CON LOS DATOS

Una vez el programa recibe los archivos de texto en el formato indicado en el apartado anterior, estos se han de limpiar y optimizar, para la mayor eficiencia, rendimiento del programa. De esta manera podremos sacar datos más precisos, reduciendo costes de tiempo.

Para llevar a cabo esta optimización, debemos dividir esta actividad en varios procesos:

- El primer proceso sería el de capitalizar todas las letras de la opinión para dejarlas todas en el mismo formato.
- El segundo proceso, y para acortar y optimizar la valoración procederemos a tokenizar las frases.
- El tercer proceso, sería el de añadir una lista de palabras de parada (Stop-Words), para que el programa omita palabras que no aporten ningún tipo de información o significado a la frase. En este paso, se le permitirá al usuario elegir una lista de palabras de parada

que se darán por defecto en el programa, o bien que el usuario sea capaz de importar su propia lista de palabras de parada.

- El cuarto, y último paso, será el de reducir las palabras de las valoraciones a las raíces de las mismas.

### 5.3 DESCRIPCIÓN DEL ALGORITMO A USAR

Nuestra idea no es la de ofrecer al usuario un único algoritmo y obligatorio, sino la de darle una lista de algoritmos donde pueda elegir, decidir y ver cual es el que más se le ajusta en cada caso. Entre los algoritmos a introducir podemos destacar los siguientes:

- Árbol de decisión (decision tree): Se construye un árbol de decisión basándose en observaciones sobre el conjunto de entrenamiento. Construye unas reglas lógicas sencillas que permiten clasificar los nuevos datos.
- Naive Bayes: Este clasificador asume que las variables predictivas son independientes. Utiliza el teorema de Bayes y requiere pocos datos de entrenamiento.
- KNN: entrenamiento mediante ejemplos cercanos en el espacio de los elementos. knn es un tipo de aprendizaje vago donde la función se aproxima solo localmente.
- Regresión logística: Es un tipo de análisis de regresión utilizado para predecir el resultado de una variable que puede tomar un número limitado de categorías, en función de las variables predictoras
- Random Forest: Algoritmo donde se basa en una serie de árboles de decisión. En un árbol de decisión, una entrada se introduce en la parte superior y baja hacia abajo, a medida que se atraviesa el árbol con los datos, al final se acumulan en conjuntos más pequeños.
- SVM: Construye un modelo capaz de predecir si un punto nuevo (cuya categoría desconocemos) pertenece a una categoría o a la otra. En este caso predecir si una nueva valoración pertenece a un grupo u otro

### 5.4 DESCRIPCIÓN DEL “CROSS VALIDATION”

Este operador es donde internamente se hace el modelo con el algoritmo seleccionado y luego se hace el testing, luego también se utiliza para volver a probar el modelo con un dataset diferente al que se utilizó para la creación del modelo. El detalle del funcionamiento particular de cada algoritmo está en otro apartado de este documento. El cross validation es uno de los principales operadores que se necesitan para este tipo de proyecto.



## 5.5 DESCRIPCIÓN DE LA INTERFAZ DE USUARIO

| Paso | Acción   |
|------|--|
| 1    | El usuario busca en su equipo la carpeta con las valoraciones clasificadas |
| 2    | El usuario indica la carpeta de valoraciones al sistema                    |
| Paso | Acción   |
| 1    | El usuario selecciona el algoritmo para clasificar las valoraciones        |
| Paso | Acción   |
| 1    | El usuario selecciona el algoritmo para clasificar las valoraciones        |

|             |   |
|-------------|---|
| 2           | El algoritmo clasifica las valoraciones   |
| 3           | El usuario hace clic en el botón "Ver valoraciones"   |
| <b>Paso</b> | <b>Acción</b>   |
| 1           | El programa muestra al usuario las opciones que puede elegir para la fase de entrenamiento. |
| 2           | El usuario indica el algoritmo que desee.   |
| <b>Paso</b> | <b>Acción</b>   |
| 1           | El programa muestra al usuario la clasificación de todas opiniones.                         |
| 2           | El usuario indica la visualización deseada.   |
| <b>Paso</b> | <b>Acción</b>   |

|   |  |
|---|--|
| 1 | El programa muestra al usuario la clasificación de todas opiniones.          |
| 2 | El usuario indica la clasificación de una o varias valoraciones en concreto. |

## 5.6 DESCRIPCIÓN DE LA SALIDA QUE GENERARÁ LA APLICACIÓN

El programa una vez clasificado las valoraciones según intérprete como buena, neutra o mala, exportará una carpeta con el nombre del proyecto y dentro de esta, tres carpetas, en función de que tipo sea cada opinión, meterá la valoración en una o en otra. Por ejemplo, si una valoración es buena, esta valoración será introducida dentro de la carpeta 'Buenas', o por el contrario si una valoración es mala, esta valoración será introducida dentro de la carpeta 'Malas'.

## 5.7 DESCRIPCIÓN DEL ENTREGABLE QUE RECIBIRÁ EL CLIENTE

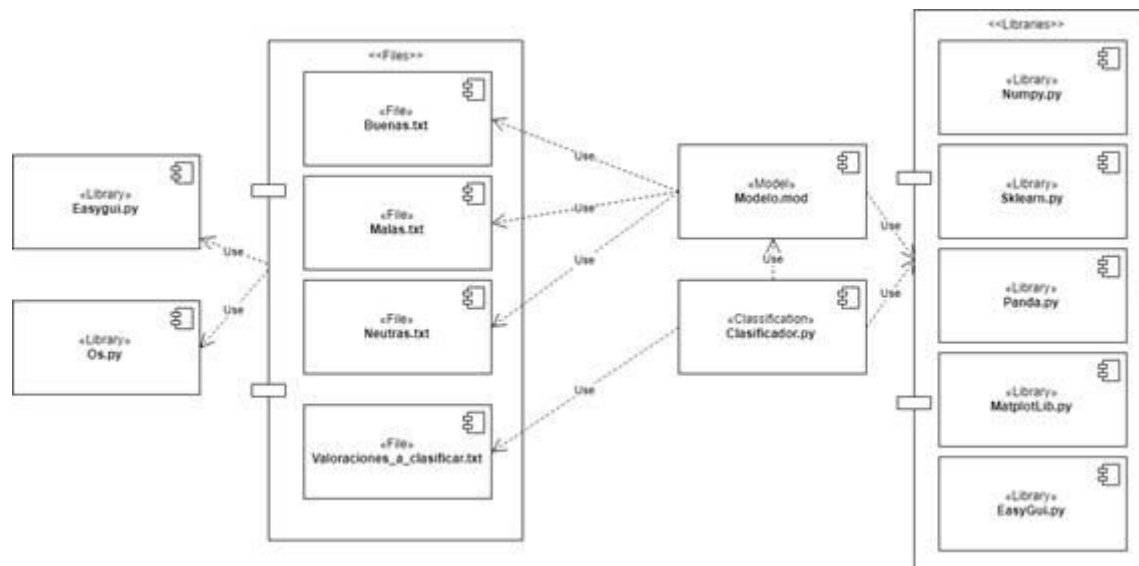
Nuestro proyecto se trata de una aplicación basada en python, un lenguaje de programación orientado a objetos que permite a los programadores o desarrolladores crear aplicaciones que pueden ser ejecutadas en cualquier equipo.

Esta aplicación será creada y diseñada pensando en ofrecer al usuario una experiencia agradable y sencilla. Además, al tratarse de una aplicación multiplataforma, el usuario podrá disfrutar de ella en casi cualquier dispositivo existente. Nuestro sistema tiene como principal objetivo ser de gran utilidad para el cliente, contando con una usabilidad intuitiva y amigable para el usuario y será multiplataforma, lo que significa que el cliente podrá ejecutar el programa con cualquier sistema operativo ya sea windows, mac o linux. Nuestra interfaz gráfica contará con pocos botones y sus funciones serán claras. Otra característica importante de nuestro proyecto será la eficacia y rapidez, esto lo conseguiremos probando una gran variedad de algoritmos y librerías, quedándonos al final con las más se ajusten a nuestras necesidades.

El cliente, recibirá el proyecto con su código fuente, pero además, y al tratarse de un programa multiplataforma, recibirá adicionalmente un ejecutable de Windows (.exe), un ejecutable de Mac OS (.app) o bien un ejecutable de linux (tar.gz), según el cliente requiera.

## 6. DISEÑO TÉCNICO PRELIMINAR.

### 6.1 DIAGRAMA DE COMPONENTES DEL SISTEMA



### 6.2 DESCRIPCIÓN DE LAS LIBRERÍAS

#### Panda

En nuestro caso, la librería Panda será utilizada para crear DataFrames y poder compactar todas las valoraciones recogidas para posteriormente, tratarlas de manera más sencilla y así crear el modelo de entrenamiento y a continuación clasificarlas correctamente. Panda es una librería de Python destinada al análisis de datos, que proporciona unas estructuras de datos flexibles y que permiten trabajar con ellos de forma eficiente y sencilla. Un DataFrame es una estructura de datos similares a las tablas de bases de datos relacionales como SQL.

#### Sickit-learn

Librería principal empleada para generar el modelo y el clasificador de nuestras valoraciones. Con esta librería, se extraen las características de las valoraciones (los pesos de las palabras), se crea el Pipeline (encargado de pre-procesamiento del texto), la conversión mediante el uso de TF- IDF y el uso de un *Cross Validation* para la creación del modelo y para mostrar después datos importantes del modelo creado como la precisión, el *recall*, la *accuracy* y la matriz de confusión.

#### Matplotlib

La librería MatPlotLib es la encargada de generar gráficos a partir de datos contenidos en listas o arrays y será utilizada para mostrar por pantalla la matriz de confusión de nuestras valoraciones.

#### Os

Librería utilizada para la lectura/escritura de ficheros de texto y su gestión. Esta librería nos permite acceder a funciones dependientes del sistema operativo como bien pueden ser la lectura/escritura de ficheros de texto y su gestión y la manipulación de la estructura de directorios.

### Numpy

Numpy es una de las librerías más importantes de Python, con ella, tenemos la posibilidad de operar con cualquier dato numérico o array. En nuestro caso particular, esta librería estará centrada en la redimensión de *arrays* para luego poder trabajar con ellos con la librería Sklearn.

### PyQT 5

La librería PyQT 5 está centrada en la creación de interfaces gráficas. Destaca por ser una librería que aporta unas interfaces gráficas fáciles de programar.

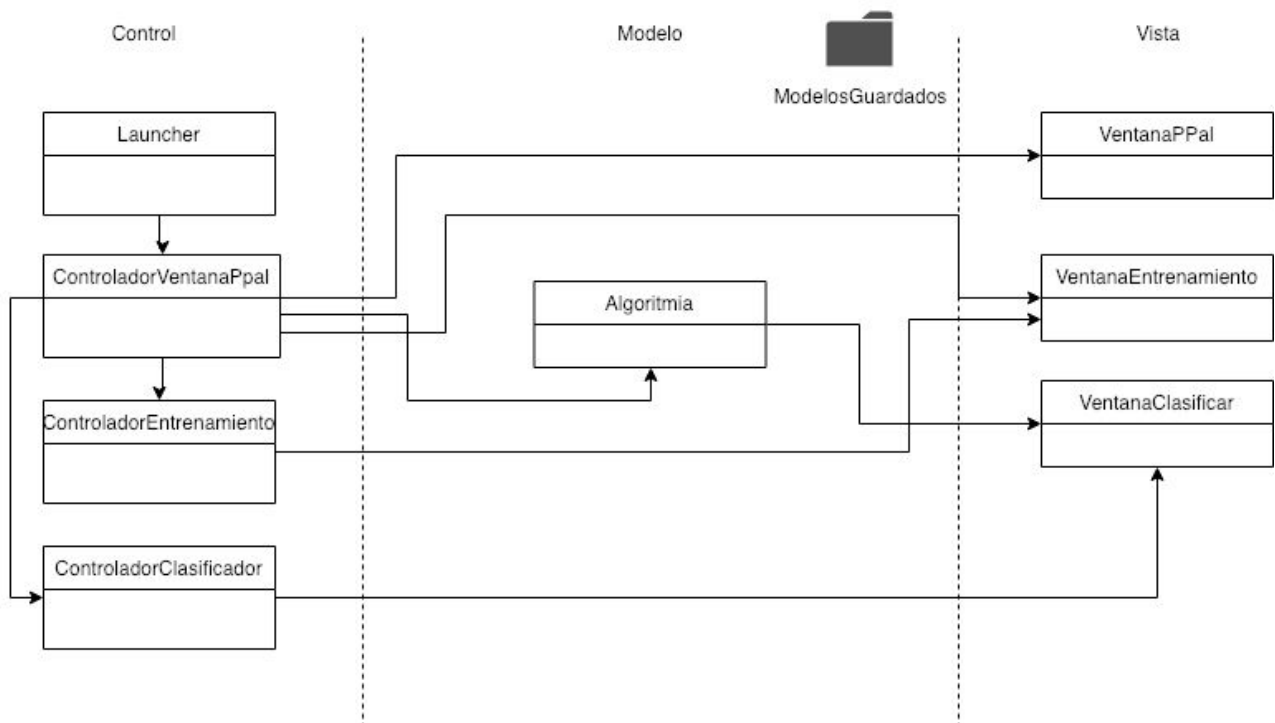
### LangDetect

La librería LangDetect es la librería que nos permite detectar el código del idioma (ej: ES) de las valoraciones sin necesidad de seleccionar el idioma manualmente.

### Pycountry

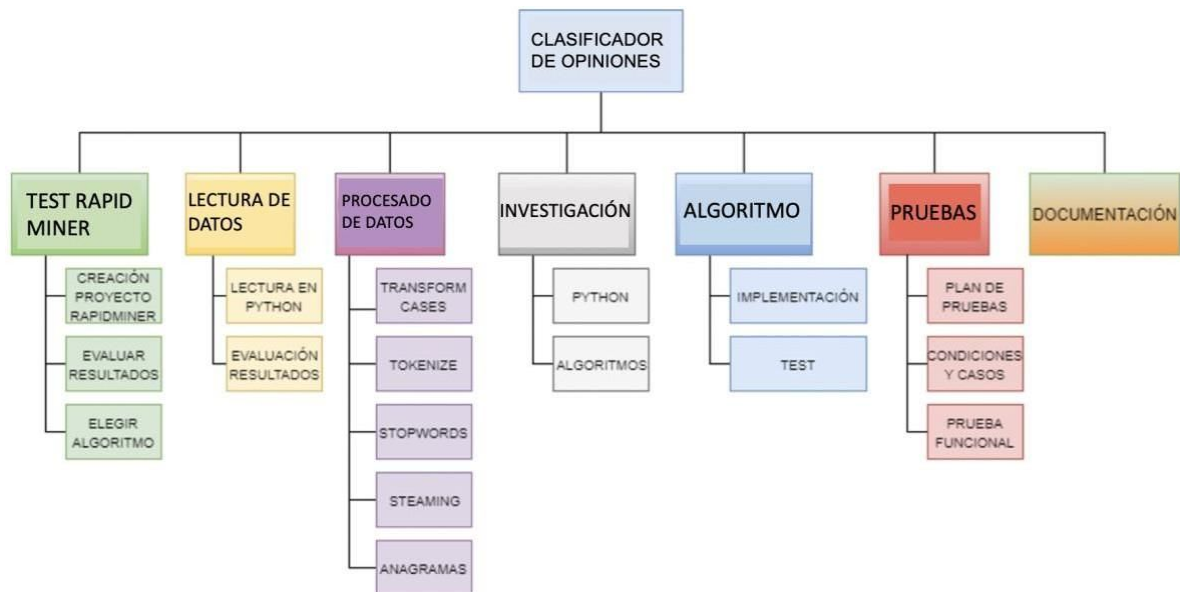
La librería Pycountry es la librería que nos permite una vez detectado código del idioma, obtener el idioma de los textos a entrenar/clasificar

## 6.3 DIAGRAMA DE CLASES



## 7. PLAN DE PROYECTO

### 7.1 DIAGRAMA EDT



### 7.2 DIAGRAMA DE GANTT

| Procesos\Semanas            | Semana 1 | Semana 2 | Semana 3 | Semana 4 | Semana 5 | Semana 6 | Semana 7 |
|-----------------------------|----------|----------|----------|----------|----------|----------|----------|
| Elicitación de requisitos   | ■        |          |          |          |          |          |          |
| Desarrollo del dossier      | ■        | ■        |          |          |          |          |          |
| Introducción a python       |          | ■        | ■        | ■        |          |          |          |
| Modelo Entrenamiento        |          |          | ■        | ■        | ■        |          |          |
| Modelo Clasificador         |          |          |          | ■        | ■        |          |          |
| Interfaz Grafica            |          |          |          |          | ■        | ■        |          |
| Cierre(Documentación final) |          |          |          |          |          | ■        | ■        |
| Extras                      |          |          |          |          |          |          | ■        |

### 7.3 PRESUPUESTO

| Presupuesto de Proyecto - Clasificación de valoraciones |                                      |   |                  |                  |         |                   |   |
|---|--------------------------------------|---|------------------|------------------|---------|-------------------|---|
| Integrantes del proyecto:                               |                                      | José Aceituno García<br>Miguel Romero García<br>Diego Gil Edo<br>Diego Abdul Massih López |                  |                  |         |                   | <b>PRESUPUESTO TOTAL</b><br><b>3.108,00 €</b> |
| Fecha de inicio:  |                                      | 10/12/18  |                  | Horas empleadas: |         | 176               |   |
| Nombre  | Elemento                             | Concepto  | Tipo de unidades | Horas/Cantidad   | Tasa    | Presupuesto       |   |
| 1 Diseño  |                                      |   |                  | 36               |         | <b>540,00 €</b>   |   |
| Front-End   | Administrador de sistemas            |   | Horas            | 12               | 15,00 € | 180,00 €          |   |
| Back-End  | Administrador de sistemas            |   | Horas            | 12               | 15,00 € | 180,00 €          |   |
| Diseño  | Administrador de sistemas            |   | Horas            | 12               | 15,00 € | 180,00 €          |   |
| 2 Ejecución   |                                      |   |                  | 140              |         | <b>2.568,00 €</b> |   |
| 2.1 Desarrollo  |                                      |   |                  | 108              |         | <b>2.160,00 €</b> |   |
| Interfaz gráfica  | Administrador de sistemas            | Tasa  | Horas            | 12               | 20,00 € | 240,00 €          |   |
| Implementación entrenamiento                            | Administrador de sistemas            | Tasa  | Horas            | 48               | 20,00 € | 960,00 €          |   |
| Implementación clasificación                            | Administrador de sistemas            | Tasa  | Horas            | 24               | 20,00 € | 480,00 €          |   |
| Funciones extras  | Administrador de sistemas            | Tasa  | Horas            | 24               | 20,00 € | 480,00 €          |   |
| 2.2 Pruebas de usuario                                  |                                      |   |                  | 32               |         | <b>408,00 €</b>   |   |
| Reuniones con el cliente                                | Clientes y Administrador de Sistemas | Tasa  | Horas            | 4                | 15,00 € | 60,00 €           |   |
| Pruebas funcionales                                     | Administrador de sistemas            | Tasa  | Horas            | 4                | 15,00 € | 60,00 €           |   |
| Reunión final con el cliente                            | Externo                              | Tasa  | Horas            | 24               | 12,00 € | 288,00 €          |   |

### 7.4 TABLA DE DESCRIPCIÓN DE FUNCIONALIDADES

| Requisitos iniciales   |   | Requisitos extra                                    |   |
|--|---|---|---|
| Interfaz Gráfica Sencilla                                    | ✓ | Visualización de los textos                         | ✗ |
| Selección de archivos de entrada y salida                    | ✓ | Selección de diferentes algoritmos de entrenamiento | ✓ |
| Generación de modelo de entrenamiento con opción de guardado | ✓ | Visualización avanzada de resultados                | ✓ |
| Ejecución de modelo de clasificador de valoraciones          | ✓ | Reclasificación de textos manualmente               | ✗ |
| Visualización de resultados                                  | ✓ | Autodetección de idioma                             | ✓ |

## 8. MANUAL DE INSTALACIÓN

Para la utilización de nuestro programa será necesario instalar Python en su versión 3.7 y PyCharm como entorno de desarrollo. Con respecto a las librerías necesarias para ejecutar el proyecto, deberemos instalar todas aquellas nombradas en el apartado 6.2 de este documento. Por otro lado, en caso de que falte alguna librería por instalarse, mediante PyCharm podremos instalarlas simplemente, haciendo clic en aquellas en las que al ejecutar el programa nos de error, para ello, haremos clic en la advertencia y seleccionaremos “instalar”. Adicionalmente se incluye un archivo denominado ‘*requeriments.txt*’, donde a través de PyCharm, si el interpretador no detecta que las librerías necesarias para hacer funcionar el programa están instaladas, le ofrecerá instalar de manera automática estas librerías.

## 9. TABLA DE PRUEBAS

### PRUEBAS VENTANA GENERAL

| ID             | TIPO DE PRUEBA   | RESULTADO ESPERADO                   | RESULTADO OBTENIDO                   |
|----------------|--|--------------------------------------|--------------------------------------|
| <b>PVG-001</b> | ACCESO A LA VENTANA ENTRENAMIENTO HACIENDO CLIC EN "ENTRENAMIENTO" | ACCESO A LA VENTANA DE ENTRENAMIENTO | ACCESO A LA VENTANA DE ENTRENAMIENTO |
| <b>PVG-002</b> | ACCESO A LA VENTANA ENTRENAMIENTO HACIENDO CLIC EN "CLASIFICACIÓN" | ACCESO A LA VENTANA DE CLASIFICACIÓN | ACCESO A LA VENTANA DE CLASIFICACIÓN |

### PRUEBAS VENTANA ENTRENAMIENTO

| ID             | TIPO DE PRUEBA  | RESULTADO ESPERADO   | RESULTADO OBTENIDO   |
|----------------|---|--|--|
| <b>PVE-001</b> | ENTRENAR SIN SELECCIONAR DIRECTORIO                               | MENSAJE DE ALERTA AVISANDO DE LA NECESIDAD DE SELECCIONAR UN DIRECTORIO CON VALORACIONES | MENSAJE DE ALERTA AVISANDO DE LA NECESIDAD DE SELECCIONAR UN DIRECTORIO CON VALORACIONES |
| <b>PVE-002</b> | ENTRENAR SIN SELECCIONAR DIRECTORIO VÁLIDO                        | MENSAJE DE ALERTA AVISANDO DE LA NECESIDAD DE SELECCIONAR UN DIRECTORIO VÁLIDO           | MENSAJE DE ALERTA AVISANDO DE LA NECESIDAD DE SELECCIONAR UN DIRECTORIO VÁLIDO           |
| <b>PVE-003</b> | ENTRENAR CON EL DIRECTORIO SELECCIONADO, SIN ALGORITMO Y/O IDIOMA | EL SISTEMA ENTRENA COGIENDO EL ALGORITMO Y/O IDIOMA POR DEFECTO                          | EL SISTEMA ENTRENA COGIENDO EL ALGORITMO Y/O IDIOMA POR DEFECTO                          |
| <b>PVE-004</b> | BOTÓN ATRÁS   | EL SISTEMA RETROCEDE A LA VENTANA INICIAL DEL PROGRAMA                                   | EL SISTEMA RETROCEDE A LA VENTANA INICIAL DEL PROGRAMA                                   |



|               |   |   |   |
|---------------|---|---|---|
| <b>PVE-05</b> | VISUALIZACIÓN<br>MATRIZ DE<br>CONFUSIÓN   | AL ENTRENAR EL<br>SISTEMA MUESTRA<br>LA MATRIZ DE<br>CONFUSIÓN                  | AL ENTRENAR EL<br>SISTEMA MUESTRA<br>LA MATRIZ DE<br>CONFUSIÓN                  |
| <b>PVE-06</b> | ENTRENAR CON UN<br>DIRECTORIO CON<br>MENOS/MÁS DE TRES<br>CARPETAS DE<br>VALORACIONES | EL SISTEMA<br>ENTRENA CON EL<br>DIRECTORIO<br>INDICADO<br>CORRECTAMENTE         | EL SISTEMA<br>ENTRENA CON EL<br>DIRECTORIO<br>INDICADO<br>CORRECTAMENTE         |
| <b>PVE-07</b> | GUARDAR<br>ENTRENAMIENTO<br>SIN INDICAR UN<br>NOMBRE                                  | EL SISTEMA NO<br>PERMITE GUARDAR<br>UN MODELO DE<br>ENTRENAMIENTO SIN<br>NOMBRE | EL SISTEMA NO<br>PERMITE GUARDAR<br>UN MODELO DE<br>ENTRENAMIENTO SIN<br>NOMBRE |
| <b>PVE-08</b> | GUARDAR<br>ENTRENAMIENTO<br>CON NOMBRE  | EL SISTEMA GUARDA<br>EL MODELO CON EL<br>NOMBRE INDICADO<br>POR EL USUARIO      | EL SISTEMA GUARDA<br>EL MODELO CON EL<br>NOMBRE INDICADO<br>POR EL USUARIO      |

#### PRUEBA VENTANA CLASIFICACIÓN

| ID             | TIPO DE PRUEBA                                     | RESULTADO<br>ESPERADO   | RESULTADO<br>OBTENIDO   |
|----------------|--|---|---|
| <b>PVC-001</b> | CLASIFICAR SIN<br>SELECCIONAR<br>DIRECTORIO        | MENSAJE DE ALERTA<br>AVISANDO DE LA<br>NECESIDAD DE<br>SELECCIONAR UN<br>DIRECTORIO CON<br>VALORACIONES | MENSAJE DE ALERTA<br>AVISANDO DE LA<br>NECESIDAD DE<br>SELECCIONAR UN<br>DIRECTORIO CON<br>VALORACIONES |
| <b>PVC-002</b> | CLASIFICAR SIN<br>SELECCIONAR<br>DIRECTORIO VÁLIDO | MENSAJE DE ALERTA<br>AVISANDO DE LA<br>NECESIDAD DE<br>SELECCIONAR UN<br>DIRECTORIO VÁLIDO              | MENSAJE DE<br>ALERTA AVISANDO<br>DE LA NECESIDAD DE<br>SELECCIONAR UN<br>DIRECTORIO VÁLIDO              |

|               |  |  |  |
|---------------|--|--|--|
| <b>PVC-03</b> | CLASIFICAR CON EL DIRECTORIO SELECCIONADO, SIN MODELO Y/O IDIOMA | EL SISTEMA CLASIFICA COGIENDO EL MODELO Y/O IDIOMA POR DEFECTO                           | EL SISTEMA CLASIFICA COGIENDO EL MODELO Y/O IDIOMA POR DEFECTO                           |
| <b>PVC-04</b> | BOTÓN ATRÁS  | EL SISTEMA RETROCEDE A LA VENTANA INICIAL DEL PROGRAMA                                   | EL SISTEMA RETROCEDE A LA VENTANA INICIAL DEL PROGRAMA                                   |
| <b>PVC-05</b> | VISUALIZACIÓN DE RESULTADOS                                      | AL CLASIFICAR EL SISTEMA MUESTRA UNA GRÁFICA DE BARRAS                                   | AL CLASIFICAR EL SISTEMA MUESTRA UNA GRÁFICA DE BARRAS                                   |
| <b>PVC-06</b> | GUARDAR CLASIFICACIÓN EN UN DIRECTORIO INDICADO POR EL USUARIO   | EL SISTEMA GUARDA LAS VALORACIONES CLASIFICADAS EN EL DIRECTORIO INDICADO POR EL USUARIO | EL SISTEMA GUARDA LAS VALORACIONES CLASIFICADAS EN EL DIRECTORIO INDICADO POR EL USUARIO |

## 10. BIBLIOGRAFÍA

[1] <https://scikit-learn.org/stable/documentation.html>

[2] <https://www.datacamp.com>

[3] <https://stackabuse.com>

[4] <https://stackoverflow.com>

[5] <https://matplotlib.org/>

[6] <https://plot.ly/>

[7] <https://docs.scipy.org/doc/>

[8] <https://github.com/Mimino666/langdetect>

[9] <https://github.com/PEAT-AI/pycountry>

[10] <https://es.wikipedia.org>

[11] <https://wiki.python.org>

[12] <http://pyqt.sourceforge.net/Docs/PyQt5/>