



Universidad Europea Madrid

LAUREATE INTERNATIONAL UNIVERSITIES

**ESCUELA DE ARQUITECTURA, INGENIERÍA Y DISEÑO
GRADO EN INGENIERÍA INFORMÁTICA**

PROYECTO DE COMPUTACIÓN III

SEGUNDA FASE DEL PROYECTO

DIEGO ABDUL MASSIH LOPEZ

ZORAN CERRILLO DEL PINO

LAURA VIZCAINO ROCA

DIEGO GIL EDO

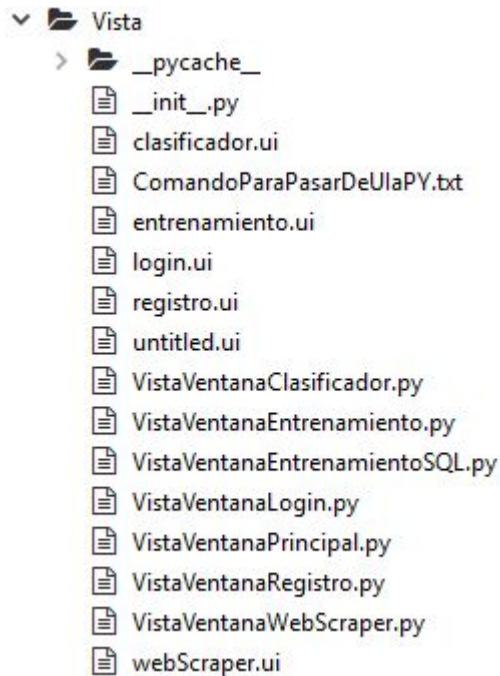
CURSO 2018-2019

Índice

Índice	2
Arquitectura general del proyecto	3
Breve manual de instalación	7
Manual de usuario	8
Aspectos destacables	14
Base de datos	14
Repositorio GitLab	15

Arquitectura general del proyecto

A continuación podemos ver cómo tenemos dividido nuestro proyecto en tres carpetas: vista, controlador y utilidades. En esta segunda fase tenemos más clases que en la primera, ahora tenemos 7 vistas:

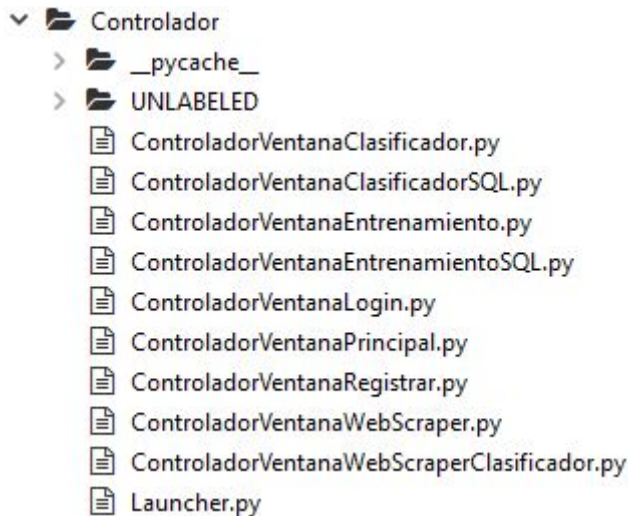


- **VistaVentanaClasificador.** En esta ventana tenemos dos opciones: Clasificar con opiniones ya existentes en la base de datos o con un WebScraper. Cuando se pulsa el botón “clasificar”, se habilitan los textos de esa lista para poder hacer doble click sobre ellos, abrirlos y ver las opiniones con su análisis de sentimiento correspondiente. Además de eso, en la parte de la derecha se muestra un gráfico con el número de valoraciones y en la parte inferior un botón para guardar la clasificación final en caso de ser necesario.
- **VistaVentanaEntrenamiento.** En esta ventana tenemos dos opciones: entrenar con opiniones ya existentes en la base de datos o con un WebScraper. Si seleccionamos la primera opción tenemos que escoger el proyecto con valoraciones que queremos entrenar y después seleccionar el algoritmo que vamos a utilizar para entrenar dichas opiniones, al pulsar el botón “entrenar” aparece una matriz de confusión en base a las valoraciones anteriores. Una vez entrenadas, tenemos la opción de guardar el resultado obtenido. Por otra parte tenemos la opción de pulsar el botón

WebScraper, que nos abrirá una ventana nueva llamada VistaVentanaWebScraper, que explicaremos más adelante.

- **VistaVentanaEntrenamientoSQL.** En esta ventana podemos seleccionar qué proyectos queremos entrenar de las desplegadas en el ComboBox. Estos proyectos fueron obtenidos anteriormente via WebScraper.
- **VistaVentanaLogin.** Es la ventana que aparece al ejecutar el programa. Existen dos roles, cliente y administrador. Dependiendo con cual entremos podremos acceder a unas funcionalidades u otras. El cliente sólo puede clasificar opiniones, mientras que el administrador accede a la VistaVentanaPrincipal.
- **VistaVentanaPrincipal.** Esta ventana es la que aparece cuando entramos como administrador. En ella nos muestra tres opciones: clasificar, entrenar o registrar un usuario nuevo.
- **VistaVentanaRegistro.** En esta ventana es donde el administrador de la aplicación podrá registrar a otros usuarios con el rol de cliente o de administrador. Simplemente deberá de rellenar los campos propuestos como el nombre, apellidos, usuario, contraseña y el rol que desee. Una vez hecho esto pulsa el botón “registrar” y el usuario quedará registrado en la base de datos.
- **VistaVentanaWebScraper.** En esta ventana tenemos que seleccionar la url en la cual vamos a buscar, nosotros actualmente tenemos 3 (booking, amazon y tripadvisor). Después de seleccionar la url deberá de introducir la url del producto u hotel del cual desee obtener las opiniones, el idioma y el número de opiniones que quiera. Una vez seleccionado todo esto, le dará al botón de obtener y se ejecutará el web scraper, de tal forma que las opiniones aparecen a tiempo real en una lista y también en el apartado de opiniones buenas, malas y totales donde se cuenta la cantidad de opiniones que hay de cada tipo. Una vez obtenidas, podemos guardarlas o borrarlas para comenzar de nuevo.

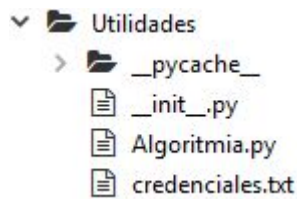
En la carpeta controlador tenemos los siguientes archivos:



- **Launcher.** En este apartado es donde definimos el main y ejecutamos el programa.
- **ControladorVentanaWebScraper.** En este controlador creamos un proyecto y hacemos el WebScraper de Booking, Amazon y TripAdvisor. Una vez hecho, tenemos el método “guardar”, el cual pregunta si queremos guardar los cambios realizados. Por otra parte, si no queremos guardar los cambios, podemos borrarlos y comenzar de nuevo.
- **ControladorVentanaRegistrar.** Aquí es donde el administrador puede añadir nuevos usuarios o clientes cubriendo los datos que se piden y seleccionando el rol.
- **ControladorVentanaPrincipal.** Este controlador maneja la ventana principal, donde decidimos si entrenar el programa, clasificar opiniones o registrar un nuevo usuario.
- **ControladorVentanaLogin.** En este controlador hacemos una consulta a la base de datos para acceder al programa como administrador o cliente.
- **ControladorVentanaEntrenamientoSQL.** Este controlador recoge las valoraciones obtenidas de un proyecto anteriormente via WebScraper y las vuelve a pasar a Entrenamiento.
- **ControladorVentanaEntrenamiento.** Este controlador tiene varios métodos pero los principales son:
 - Filechooser. Encargado de introducir el url con las valoraciones a entrenar.

- Entrenamiento. Encargado de entrenar las valoraciones con el algoritmo seleccionado.
 - Guardar. Solicita al usuario que asigne un nombre al entrenamiento realizado y lo guarda en la base de datos.
 - ConfigurarFiguraGrafico. Método encargado de dibujar las gráficas con los datos obtenidos.
 - HabilitarOpcionesCheckBox.Método encargado de habilitar o deshabilitar el ComboBox de idiomas dependiendo del idioma en el que se encuentre.
 - Diálogo. Método que se encarga de crear un diálogo de tipo informativo.
- **ControladorVentanaClasificador.** Igual que en el controlador anterior, explicaremos los métodos más importantes:
 - Filechooser. Encargado de introducir el url con las valoraciones a clasificar.
 - RellenarModelo. Agrega al ComboBox los modelos disponibles.
 - AnalisisSentimiento. Especifica en cada texto la polaridad que tienen sus opiniones.
 - Clasificar. Método encargado de clasificar las opiniones seleccionadas.
 - ConfigurarListaParaGrafico. Genera los datos necesarios para representar las gráficas con sus opiniones.
 - ConfigurarGrafico. Dibuja las gráficas con los datos obtenidos anteriormente.
 - Guardar. Método creado para guardar los textos ya clasificados.
- **ControladorVentanaClasificadorSQL.** Este controlador clasifica las opiniones que están guardadas en la base de datos y muestra la nota y el texto de cada una.
- **ControladorVentanaWebScraperClasificador.** Aquí se ejecutan las funciones para clasificar las opiniones obtenidas mediante un WebScraper. Seleccionamos una página y en base a eso, se van contando las opiniones buenas, malas y el total a tiempo real. Cuando se finaliza el número indicado de opiniones aparece un mensaje de aviso. Si el usuario lo desea, puede guardar la clasificación obtenida o eliminarla.

En la carpeta utilidades tenemos la clase Algoritmia.py.



Lo que hace esta clase es entrenar datos según un algoritmo seleccionado o detectar en qué idioma están las opiniones obtenidas. Después clasifica y procesa esos datos.

Breve manual de instalación

Para la parte de WebScraper necesitamos instalar una serie de herramientas para que funcione correctamente.

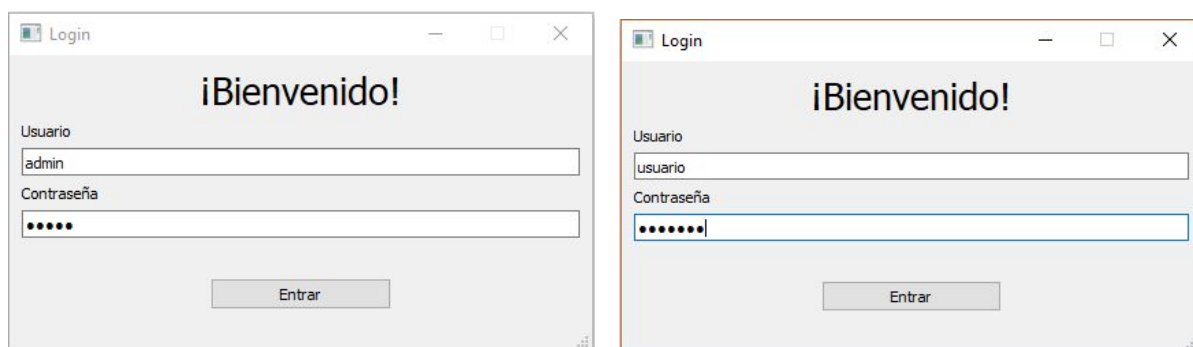
Boto3: Sirve para usar AWS de forma rápida y eficiente, es un AWS SDK para Python. Boto3 facilita la integración de su aplicación, biblioteca o script de Python con los servicios de AWS, incluidos Amazon S3, Amazon EC2, Amazon DynamoDB etc... En nuestro caso, lo usamos para guardar los modelos en Amazon Web Services. Los modelos que utilizamos en nuestro proyecto se guardan su nombre en la base de datos en una tabla llamada "modelos" donde se referencian con la base de datos no relacional de Amazon.

Beautiful Soup: es una biblioteca de Python para analizar documentos HTML. Esta biblioteca crea un árbol con todos los elementos del documento y puede ser utilizado para extraer información. Por lo tanto, esta biblioteca es útil para realizar web scraping. La hemos utilizado para llevar a cabo los nuestros.

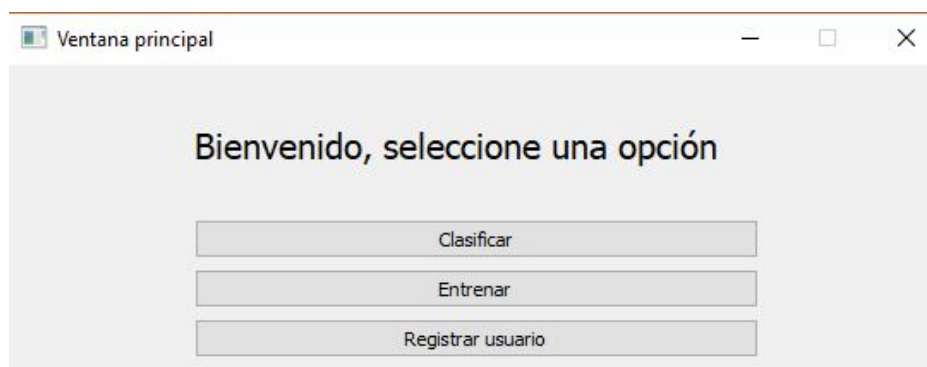
TextBlob: es una biblioteca Python que sirve para procesar datos textuales. Proporciona una API simple para sumergirse en tareas comunes de procesamiento de lenguaje natural (NLP), como etiquetado de parte del discurso, extracción de frases nominales, análisis de sentimientos, clasificación, traducción etc... Nosotros la usamos para realizar el análisis de sentimientos y obtener la polaridad de las opiniones con las que trabajamos.

Manual de usuario

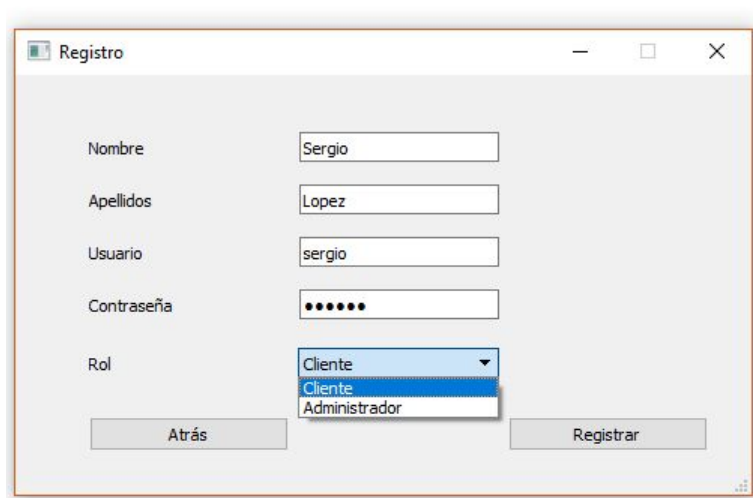
Al iniciar el programa aparece una ventana de Login donde podemos entrar como administrador o cliente. Para entrar como administrador el usuario es **admin** y la contraseña **admin**. Si queremos entrar como clientes el usuario es **usuario** y la contraseña **usuario**.



Al entrar como administrador aparece una nueva ventana para seleccionar una de las siguientes opciones:



El administrador puede registrar usuarios. Aquí vemos la ventana donde se pueden añadir nuevos usuario o clientes, según el rol que se elija. El administrador debe cubrir los campos que se muestran a continuación y pulsar el botón "Registrar". Este nuevo usuario se guardará en la base de datos y ya puede acceder al programa con unas opciones u otras, dependiendo del rol seleccionado.



Registro

Nombre: Sergio

Apellidos: Lopez

Usuario: sergio

Contraseña: •••••

Rol: Cliente (dropdown menu showing Cliente and Administrador)

Atrás Registrar

Si seguimos con el rol de administrador y en la ventana principal pulsamos “entrenar” aparece una ventana, donde tenemos dos opciones: hacer un WebScraper o entrenar con unas opiniones existentes en la base de datos. Al pulsar el botón WebScraper de la ventana anterior accedemos a una nueva ventana donde tenemos un ComboBox para elegir una página (booking, amazon o tripadvisor) de donde queremos obtener las opiniones. A continuación nos pide indicar la URL de dicha página, aquí dejamos un ejemplo de cada uno con los que hemos hecho pruebas:

- https://www.tripadvisor.es/Restaurant_Review-g187849-d2048527-Reviews-Nerino_Dieci_Trattoria-Milan_Lombardy.html
- https://www.amazon.es/SPARIN-Protector-Pantalla-iPhone-Definicion/dp/B078N9L854/ref=sr_1_5?_mk_es_ES=%C3%85M%C3%85%C5%BD%C3%95%C3%91&crd=1LP7HLWCXILUL&keywords=iphone+7+cristal+templado&qid=1554711168&s=gateway&srefix=iphone+7+cristal%2Caps%2C155&sr=8-5
- <https://www.booking.com/hotel/ae/premier-inn-dubai-silicon-oasis.es.html?label=gen173nr-1BEgdyZXZpZXdzKIICOOgHSDNYBGipAYgBAZgBCrgBF8gBD9gBAegBAYgCAagCA7gCIISj5AXAAgE:sid=01d9e8adb79a678bb7b88e9199cc639c>

También debemos seleccionar el idioma en el que queremos que se recuperen las opiniones. En este caso puede ser inglés o español(solo Booking) el resto las toma en español. Debemos de indicar el número máximo de opiniones de cada tipo que queremos entrenar. Podemos elegir entre 75, 150, 225, 300, 375, 450, 525, 600, 675, 2700 o 5400.

Cuando pulsamos el botón “OBTENER” nos pide que insertemos un nombre para generar un nuevo proyecto.

The screenshot shows a web scraper application interface. At the top, there's a section titled "Seleccione una página para obtener opiniones:" with a dropdown menu showing "https://www.booking.com". Below this, there's a "URL:" label and a text input field containing a long alphanumeric string. To the right of the URL field, there are three input fields for "Número de opiniones Buenas", "Número de opiniones Malas", and "Número de opiniones Total", each with a placeholder text "Aquí te avisaremos cuando terminemos". Below the URL field, there's a "Seleccione el idioma en que desea recuperar las opiniones:" dropdown menu showing "Ingles". Below that, there's a label "Ingrese el maximo de opiniones que desee:" and a dropdown menu showing "75". In the center, a modal dialog titled "Guardar" is open, asking "Introduzca el nombre del proyecto a generar:" with a text input field containing "prueba1" and "OK" and "Cancel" buttons. At the bottom, there are three buttons: "Atrás", "OBTENER", and "Guardar".

Después de esto, se ven todas las opiniones con su correspondiente nota y en la esquina superior derecha aparece un contador de las opiniones buenas, malas y totales. Como en este caso le hemos indicado 75 como máximo, el programa se detendrá al alcanzar ese número de opiniones buenas o malas, o en su defecto, al entrenar el número total de opiniones en caso de ser inferior a 75. Se puede observar cómo el campo de la derecha cambia su texto una vez estén todas las opiniones obtenidas.

Por último, podemos guardar el resultado en la base de datos o borrarlo y comenzar de nuevo. Si lo guardamos vemos que se inserta correctamente en la base de datos.

Seleccione una página para obtener opiniones:

URL:

Seleccione el idioma en que desea recuperar las opiniones:

Ingrese el maximo de opiniones que desee:

Número de opiniones Buenas:

Número de opiniones Malas:

Número de opiniones Total:

WebScraper

8.8 Welcoming reception , clean comforting rooms . Nice pup . Qu

8.8 Quite distant from transit stations . Pool is small . Poor TV . Som

8.3 Perfect service from beautiful Jane on reception . She solved my problems very quickly . Thanks for Jane . It was very quick very clear and very understandable .

8.8 If you are visiting Dragon mart , global village , outlet mall , this is the best lication . I was travelling with may family and they gave me connecting rooms . Jane a

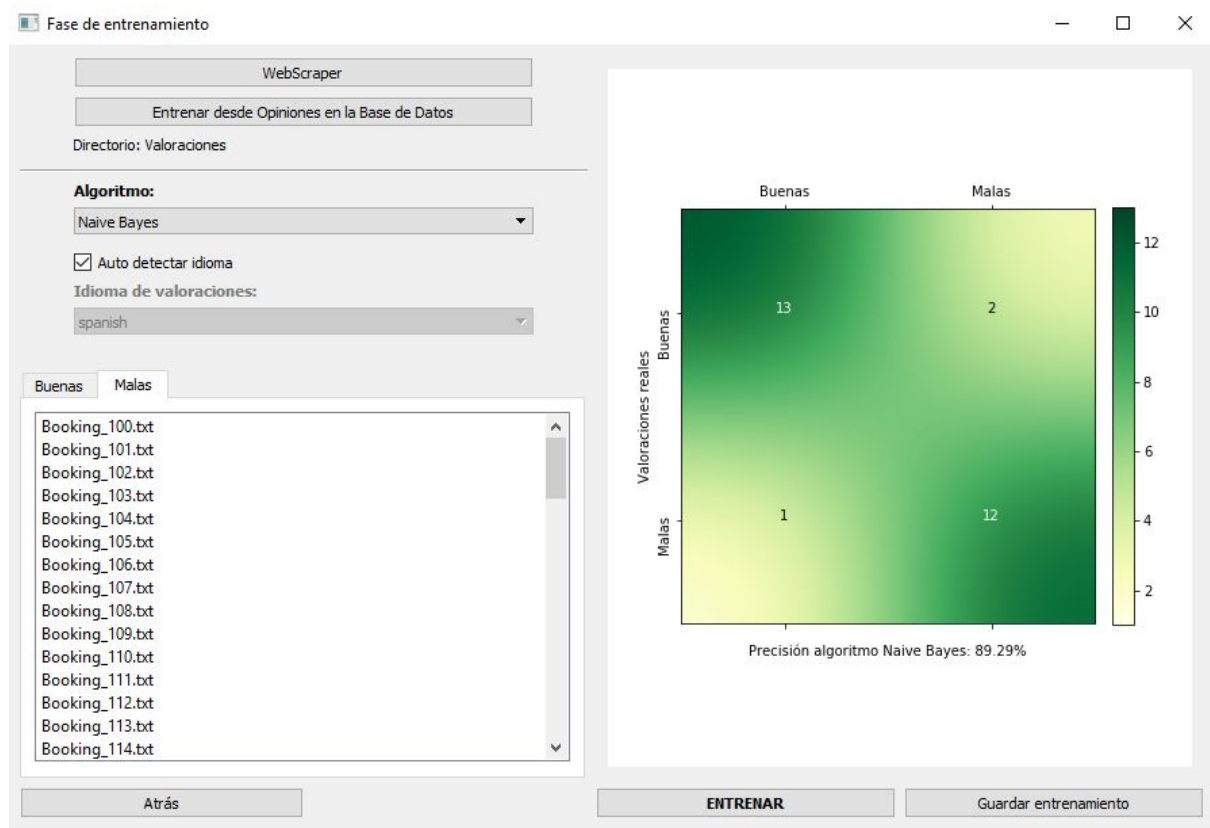
8.8 Nothing

ID_Proyecto	Nombre	ID_Usuario
7	EvidenciasEntregaWebScraper	0
21	pruebafinal	0
22	a	0
23	prueba	0
24	a	0

Por otra parte, en la ventana de entrenamiento tenemos también la opción de entrenar opiniones desde la base de datos con los proyectos guardados. Si seleccionamos esa opción nos redirige a una ventana donde podemos elegir qué proyecto entrenar, sin tener que hacer otra vez el WebScraper.

Proyectos Guardados

Después de elegir si queremos hacer WebScraper o entrenar con la base de datos, volvemos a la ventana de entrenamiento y vemos que más abajo tenemos un ComboBox donde podemos elegir el algoritmo para entrenar y más abajo el idioma, el cual está en español por defecto pero podemos cambiarlo. Una vez seleccionado todo esto, pulsamos el botón “ENTRENAR” y aparecerá a la izquierda una lista de archivos con dos pestañas (opiniones buenas y malas) con el texto y las valoraciones de cada una. A la derecha está la matriz de confusión que cambia según el algoritmo elegido. Una vez realizado el entrenamiento, podemos guardarlo presionando el botón inferior derecho y asignándole un nombre.



Por otra parte, el administrador también puede clasificar opiniones pero en el caso del cliente sólo puede acceder a la ventana de clasificar. Es una ventana muy similar a la de entrenamiento ya que también hay que hacer el WebScraper, seleccionar el idioma y el modelo de entrenamiento que está guardado en la base de datos.

ID_Modelo	Nombre
11	amazonabdul
12	amazonbayes
14	amazonregresion
15	amazonarbol
17	bookingregresion
18	booking2hoteles
19	lauritamezcla
20	modelocon3paginas
21	prueba
22	evidenciasentregawebscraper
23	booking
24	prueba1

Cuando pulsamos clasificar se vuelven a mostrar los archivos divididos entre opiniones buenas o malas, la matriz de confusión y se habilita el botón de “Guardar clasificación”.

Fase de clasificación

WebScraper

Seleccione Opiniones Recuperadas en Proyectos Anteriores

Directorio:

Modelo de entrenamiento:

amazonabdul

☒ Auto detectar idioma

Idioma valoraciones

spanish

Textos a Valorar

Atrás

CLASIFICAR

Guardar clasificación

Si en este caso pulsamos la opción de “Seleccionar opciones recuperadas de proyectos anteriores” podemos ver que toma los datos de la base de datos de proyectos ya clasificados y guardados.

Proyectos Guardados

EvidenciasWebScraperClasificador

EvidenciasWebScraperClasificador
Pruebas
p

Atrás
Seleccionar

ID_ProyectoClasificacion	Nombre	ID_Usuario
54	EvidenciasWebScraperClasificador	0
57	Pruebas	0
59	p	0

Aspectos destacables

- **Boto3** - Boto3 es el SDK de Amazon Web Services (AWS) para Python. Permite a los desarrolladores de Python crear, configurar y administrar servicios de AWS, como EC2 y S3. Boto proporciona una API fácil de usar y orientada a objetos, así como un acceso de bajo nivel a los servicios de AWS.
- **Entrenar bbdd** - Nuestro proyecto puede entrenar con opiniones obtenidas mediante un WebScraper o con opiniones recogidas de la base de datos.
- **Entrenamiento mezclado** - Nuestro proyecto tiene la opción de entrenar con las opiniones de una página y una vez haya acabado, puede entrenar con las de otra página y se mezclarán. Sin límite de páginas web.

Base de datos

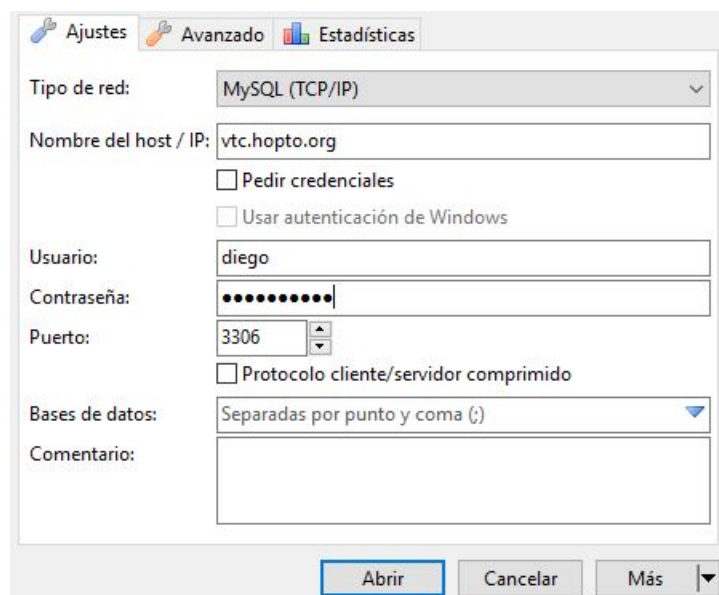
Para poder acceder a nuestra base de datos hay que poner los siguientes datos:

Nosotros hemos utilizado HeidiSQL para toda la gestión de la base de datos.

Nombre del host: vtc.hopto.org

Usuario: diego

Contraseña: Galicia96.



The image shows the 'Ajustes' (Settings) tab of the HeidiSQL application. The 'Tipo de red' (Network type) is set to 'MySQL (TCP/IP)'. The 'Nombre del host / IP' (Host / IP) field contains 'vtc.hopto.org'. There are two unchecked checkboxes: 'Pedir credenciales' (Ask for credentials) and 'Usar autenticación de Windows' (Use Windows authentication). The 'Usuario' (User) field contains 'diego'. The 'Contraseña' (Password) field is masked with dots. The 'Puerto' (Port) field is set to '3306'. There is an unchecked checkbox for 'Protocolo cliente/servidor comprimido' (Compressed client/server protocol). The 'Bases de datos' (Databases) dropdown is set to 'Separadas por punto y coma (;)' (Separated by semicolon). The 'Comentario' (Comment) field is empty. At the bottom, there are three buttons: 'Abrir' (Open), 'Cancelar' (Cancel), and 'Más' (More) with a dropdown arrow.

Repositorio GitLab

Al no poder crear un repositorio en GitHub para 5 miembros de forma privada, hemos decidido crear un repositorio en GitLab donde se necesita el siguiente enlace para acceder:

<https://gitlab.com/diegoabdul/proyectoComputacionII.git>