

Módulos Python avanzados

Scikit-learn (Scouting de un ojeador de futuras leyendas del futbol)

¿De que se trata el modulo que he seleccionado?

He elegido Scikit-learn para esta práctica porque me apetecía probar algo de Inteligencia Artificial. Es la librería que usa casi todo el mundo para temas de Machine Learning con Python y básicamente sirve para que el ordenador aprenda de datos que ya tenemos y empiece a sacar conclusiones lógicas por su cuenta.

Como al buscar información siempre me sacaba el mismo ejemplo de clasificar flores que sale en todos los manuales de Internet. Quería hacer algo que me gustara mas así que he montado un sistema de buscar información(scouting) para un ojeador de fútbol que solo busca posibles leyendas futuras o balones de oro ,campeones mundiales cosas así .

La idea es pasarle estadísticas (goles, asistencias, pases y todo eso) y que la IA nos diga quién tiene más papeletas para ser el futuro mejor fichaje para el equipo separando por posición de juego salvo el portero, que no lo estamos teniendo en cuenta dado a que sus estadísticas son totalmente diferentes. Lo bueno de Scikit-learn es que en cuanto pillas cómo se entrena el modelo te sirve para predecir prácticamente lo que quieras.

¿Por qué este módulo?

Elegí Scikit-learn porque es la herramienta estándar en la industria y me apetecía ver cómo de difícil era entrenar un modelo desde cero y también porque las posibilidades que tienes para predecir cualquier cosa son infinitas.

¿Por qué uso sklearn.ensemble?

Uso .ensemble porque me gusta ir directo a lo que mi programa necesita. Al cargar solo lo que voy a usar, evito meter herramientas innecesarias y mantengo el código ligero y ordenado. Elijo este módulo porque es donde están los modelos más potentes, esos que de verdad marcan la diferencia en los resultados; así consigo que todo funcione de forma más rápida, limpia y eficiente.

Instalación y Configuración

Antes de nada hay que tener Python instalado. Si usas Windows, vete a su web oficial, bájalo esto es muy importante, marca la casilla de "Add Python to PATH" al principio de la instalación. Si se te olvida eso, luego la consola no sabra donde buscar y no hace nada. En Mac o Linux es más fácil, abres la terminal y tiras de comandos como `>brew install python(MAC)` o `>sudo apt install python3.(Linux)`

Para instalar la librería (**Scikit-learn**) hay varias formas:

PyCharm: Entrás a New Project, en el campo donde asignas Interpreter type asignas Custom environment y le das al símbolo del +. Ahí escribes scikit-learn y le das a instalar o si ya tienes el proyecto creado entras en Settings, buscas Python Interpreter y le das al símbolo del +. Ahí escribes scikit-learn y le das a instalar.(yo he usado el PyCharm) o escribes el import en el código colocas el ratón sobre la palabra marcada(sera el modulo en cuestion) y te salta un aviso sobre acciones que puedes hacer una de ellas sera instalar le pinchas y dejas que termine de instalar muy importante, dejar que termine de instalar dado que si realizas alguna acción un que sea involuntaria tienes muchas posibilidades de romper la base de datos interna de la aplicación, he puesto varias soluciones posibles y una de ellas por si no funciona nada como ultimo recurso.

VS Code: No hay un botón de instalar como en otros programas, así que hay que hacerlo así: Primero vas al icono de los cuadraditos de la izquierda (Extensiones) y buscas la de Python de Microsoft. Hay que instalarla para que el editor entienda el código. Luego, arriba en el menú, le das a Terminal > New Terminal. Se abrirá una ventanita abajo. Ahí abajo solo tienes que escribir `>pip install scikit-learn` y darle al intro, Verás que empiezan a salir un montón de letras y barras de carga cuando termine y te deje escribir otra vez, ya está listo.

Por **consola:** Si prefieres la terminal, solo hay que poner `>pip install scikit-learn` en Windows o `>pip3 install scikit-learn` en Mac o `>sudo pip3 install scikit-learn` en Linux.

Para verificar si la instalación salió bien se usa el comando `>pip show scikit-learn`. (en cualquier SO) Si te aparece la versión y la descripción de la librería. Ya está instalado.

Posibles fallos típicos:

Lios con el nombre: Un fallo muy tonto al principio es buscar la librería como `sklearn`. En los menús de instalación hay que buscarla como `scikit-learn` pero luego dentro del código para que funcione escribimos `sklearn`.

Error a la hora de la importación: Si pones `from sklearn import` a secas, te va a dar un error de `ImportError`. Esto pasa porque la librería está organizada por familias y tienes que decirle exactamente qué quieres. Para mi proyecto la forma correcta es: `from sklearn.ensemble import RandomForestClassifier`.

Fallo a la hora de instalarlo por el pip o del IDE que estés usando no se pueda conectar correctamente con el servidor para descargarlo existen varias formas de arreglarlo:

-Forzar la instalación por terminal

A veces las interfaces gráficas fallan. Abre tu terminal (CMD, PowerShell o Terminal de VS Code) y escribe directamente:

```
>pip install scikit-learn
```

-Actualizar las herramientas de instalación

Es muy probable que tu instalador tenga problemas para comunicarse con los servidores actuales. Ejecuta esto:

```
>python -m pip install --upgrade pip setuptools wheel
```

- Problemas de conexión o Proxy

Si estás en una red de oficina, universidad o usas una VPN, es posible que el tráfico esté bloqueado. Puedes intentar saltarte la verificación de host de esta manera:

```
>pip install scikit-learn --trusted-host pypi.org --trusted-host files.pythonhosted.org
```

-Limpiar el caché de PyCharm :

PyCharm guarda una base de datos de los esqueletos de los módulos. Si esta se corrompe, verás errores extraños.

Ve al menú superior: File > Invalidate Caches...

Marca todas las casillas (especialmente Clear file system cache and Local History).

Haz clic en Invalidate and Restart.

¿Por que puede pasar ese fallo en el pip o IDE?

Conexión inestable: Tu internet parpadeó justo cuando PyCharm intentaba leer la descripción.

Certificados SSL: Tu Python no puede verificar que la conexión con PyPI sea segura.

Caché corrupta: pip tiene guardada una versión rota de la información del paquete.

Nota: Si después de instalarlo te dice que el comando no se reconoce, asegúrate de que Python esté en las Variables de Entorno (PATH) de tu sistema o sino como ultimo recurso deberas de desinstalar y borrar todos los datos que tengas de ese pycharm y volver a instalarte el pycharm.

Nota: Es recomendable tener instalado también NumPy, que es el "motor matemático" que usa Scikit-learn para procesar las estadísticas de nuestros jugadores. Normalmente se instala solo, pero si quieres asegurarte, puedes poner: >pip install scikit-learn numpy.

Código en : [Github](#)

Web oficial de [Scikit-learn](#).

Francisco Lozano Neira
DAM2