Dokumentasi Project YOUTUBE VIDEO Sentimen Analysis via Comment Section

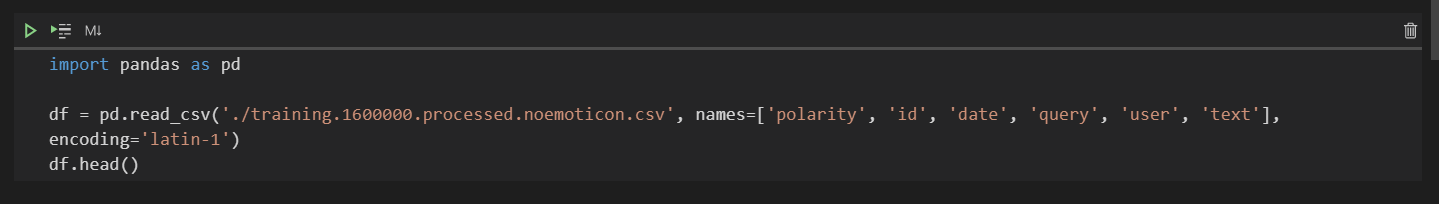
1. Cleaning dataset Sentiment140

Libary yang digunakan :

* Pandas yang digunakan untuk membuka dataset Sentiment140 dari bentuk csv ke dalam pandas DataFrame
* Pprint yang digunakan untuk mengeluarkan output
* Matplotlib yang digunakan untuk plotting sebaran data sebelum dilakukan cleaning
* BeautifulSoup4 yang digunakan untuk melakukan cleaning menggunakan lxml parser untuk menghilangkan HTML5 decode seperti ‘&quo’ dsb,
* NLTK untuk melakukan tokenizer
* Regex untuk menghilangkan mention, URL yang terdapat pada dataset Sentiment140

Coding :

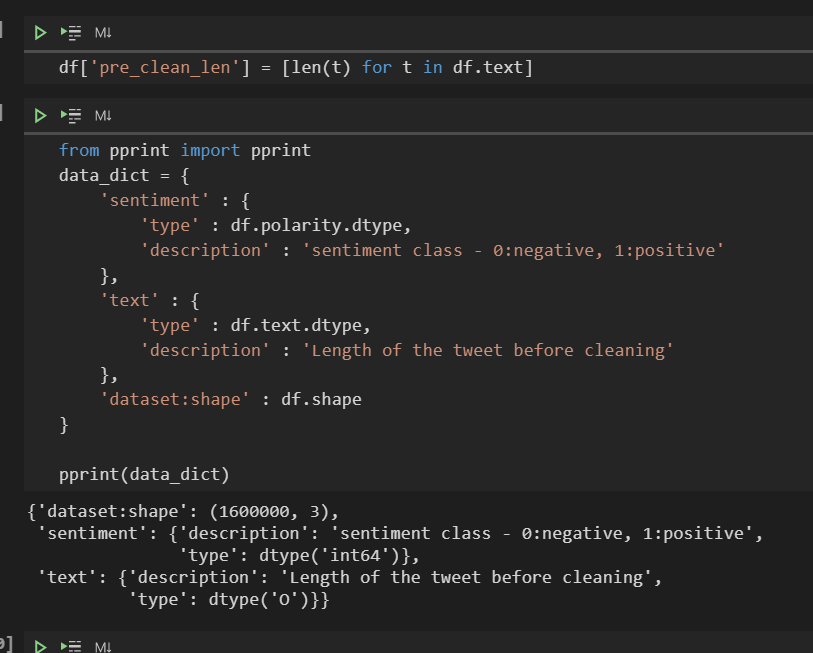
* Proses pengambilan data dan kolom yang terpakai serta melakukan encoding berdasarkan format latin-1



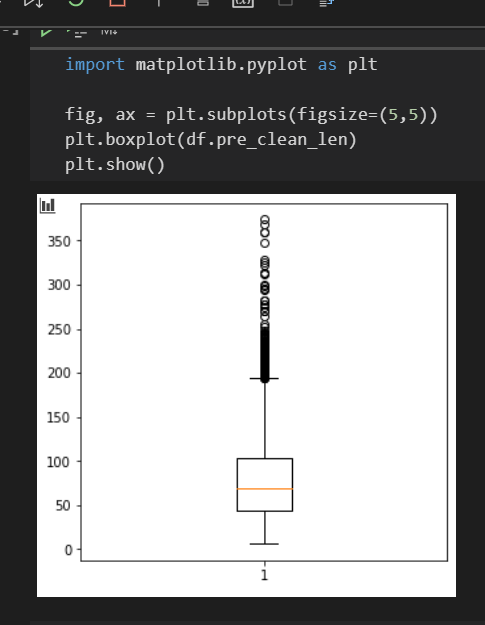
* Menghapus kolom kolom yang tidak dibutuhkan



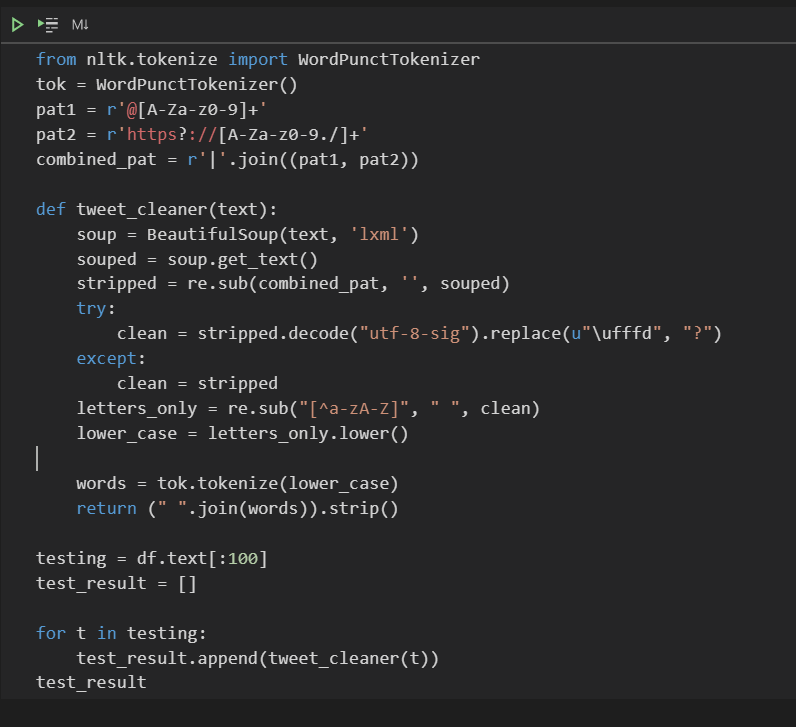
* Melakukan print detail beserta informasi tentang dataframe



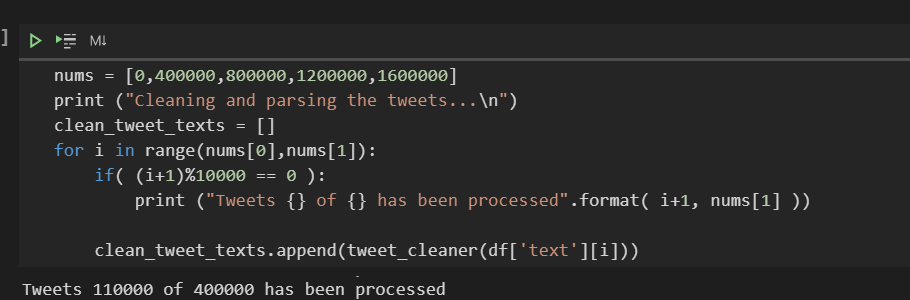
* Melakukan plotting terhadap data sebelum di cleaning



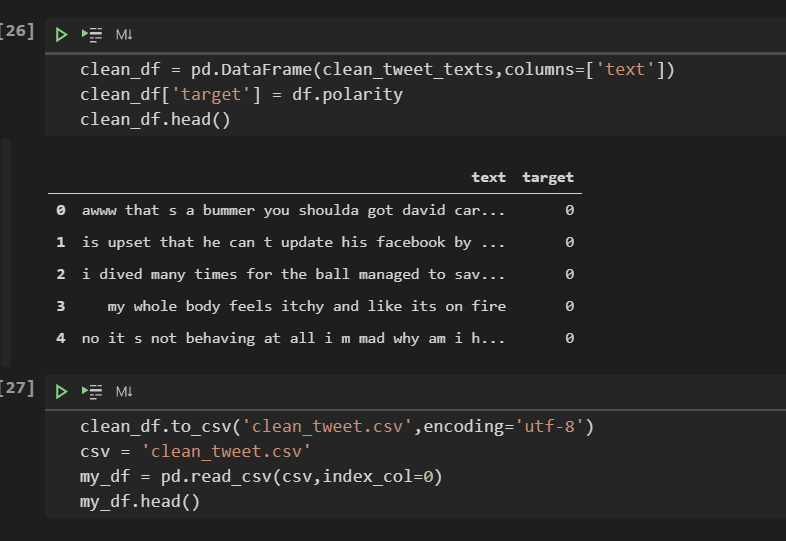
* Membuat fungsi yang berguna untuk menghilangkan HTML5 decoder, mention, URL, dsb nya lalu menampilkan beberapa data sample yang di cleaning



* Melakukan cleaning dan memasukkan tweet yang sudah dibersihkan ke dalam list yang baru



* Membuat pandas dataframe dari hasil cleaning dan mengexport data tersebut menjadi file csv yang baru.



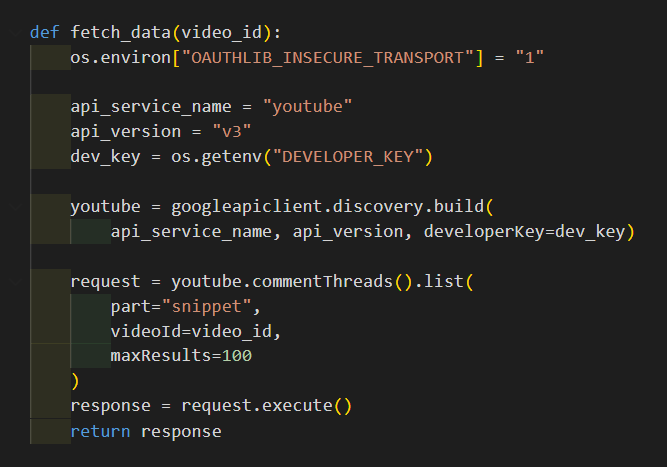
1. Mengambil Komentar menggunakan YouTube API

Library yang digunakan adalah sebagai berikut :

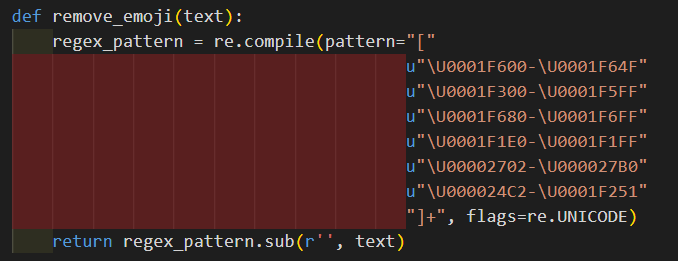
* Googleapiclient.sicoverty
* Csv
* Re
* Donev
* Deep\_translator
* Os

Berikut adalah fungsi fungsi pada program pengambil data dari YouTube API

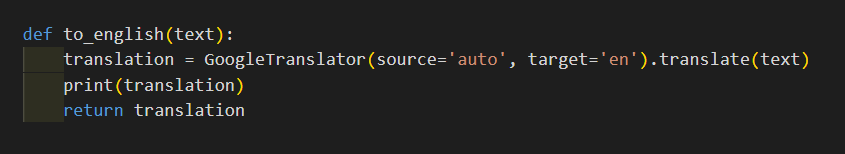
* Fungsi untuk mengambil data berdasarkan video id



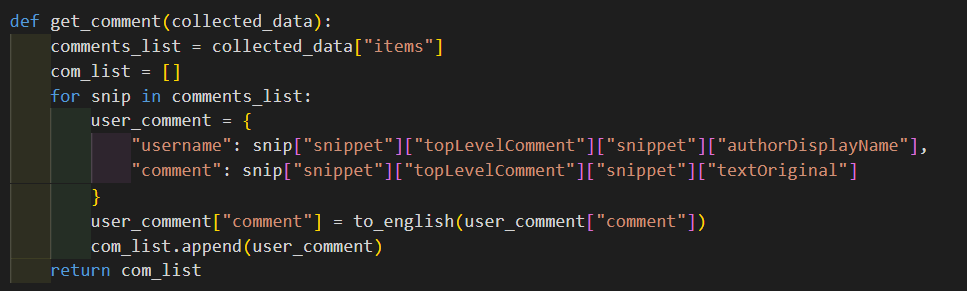
* Fungsi untuk menghapus emoji yang berada pada komentar



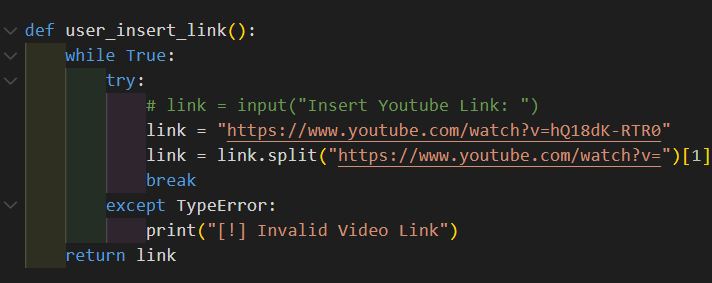
* Fungsi untuk menterjemahkan text ke dalam bahasa inggris



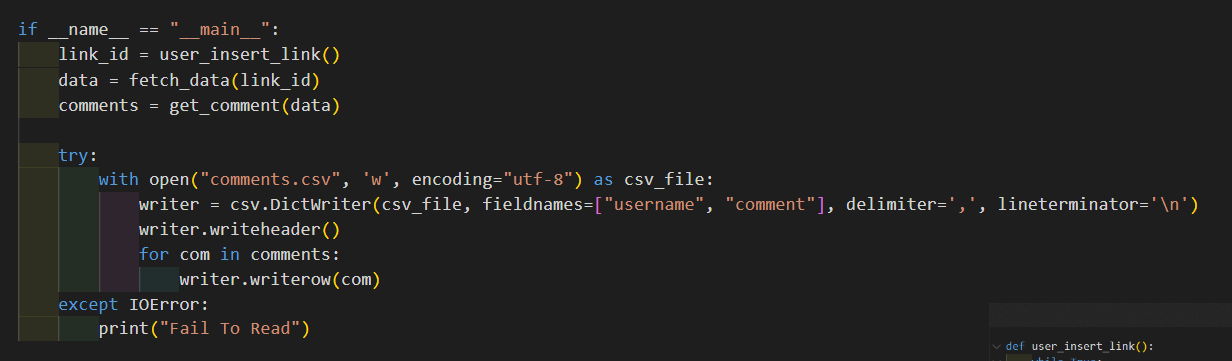
* Fungsi untuk mengambil komentar dari video YouTube



* Fungsi untuk meminta user memasukkan link video yang ingin diambil komentarnya



* Main program yang akan dijalankan pertama kali



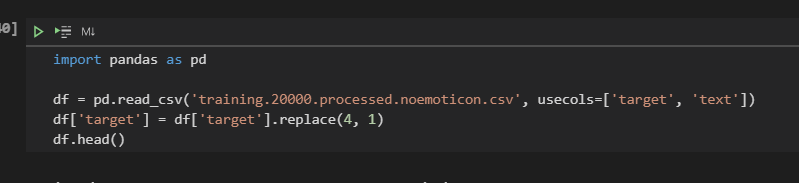
1. Training dan testing menggunakan XGBoost

Library yang dipakai :

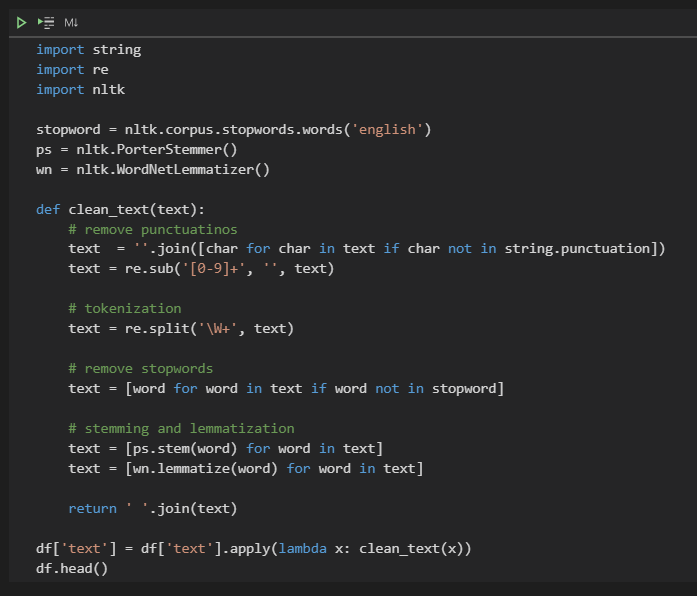
* Pandas
* Matplotlib
* Seaborn
* Sklearn
* Xgboost untuk training
* Numpy
* String
* Re
* Ntlk

Berikut adalah fungsi fungsi yang ada pada codingan :

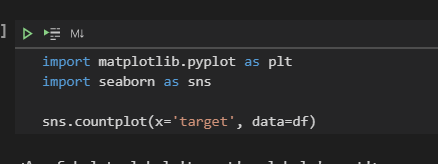
* Pengambilan dataset dari csv



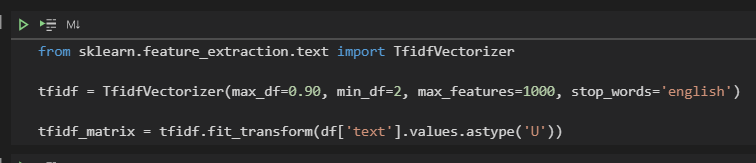
* Function cleaning



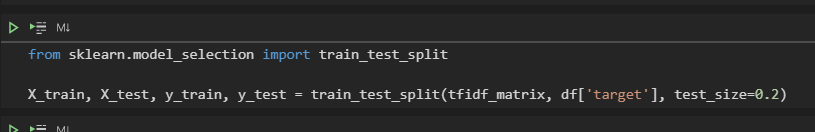
* Plotting dataset



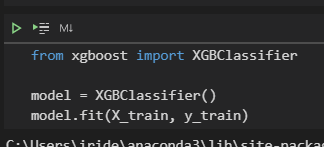
* Melakukan vectorizer



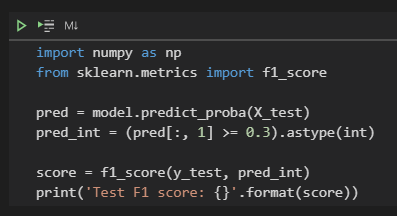
* Split test sama train



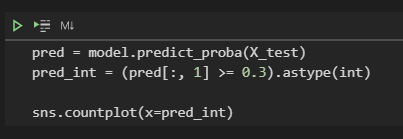
* Fit data train ke dalam XGBoost



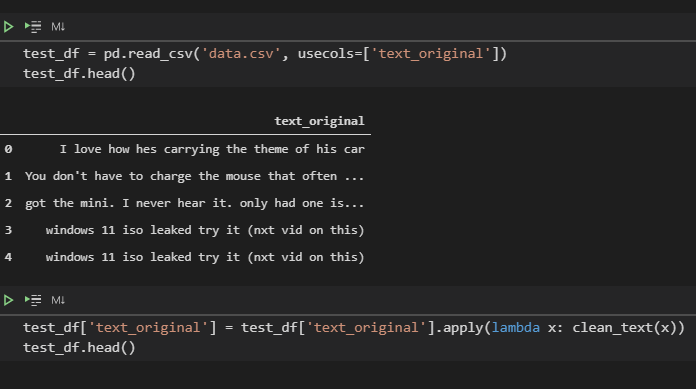
* Melakukan prediksi terhadap model



* Plotting hasil prediksi



* Pengambilan dataset dari YouTube API dan melakukan cleaning



* Dilakukan prediksi dan plotting dari hasil prediksi

