

README

Course Name : Social Media Data Mining

Project Name : TweeTrimony

Project By : Abhishek, Akshay, Akshay, Ekta, Jaspreet, Abhuranthu, Priya

Submitted on : May 10th 2022

Step 1 : Importing Header Files

Please import below header files

Format : **pip install <header_file_name>**

1. import textblob as textblob
2. import pymongo
3. import pymongo(srv) - version “3.12.1”
4. import tweepy
5. import time
6. import json
7. import regex as re
8. import pandas as pd
9. from urllib.request import urlopen

Step 2 : Please download the files present in the zip folder

File 1 : Common_Surnames_Census_2000.csv (Source : Internet)

This above file contains all the Last names present in the USA as per Census report of 2000

File 2 : SSA_Names_DB.xlsx (Source : Internet)

This above file contains all the first names present in the USA.

Note : Please download these files to a folder and please provide the path of the above files in the function **valid_users()**

Step 3 : Twitter API keys : function oauth()

Twitter API keys of all the members are already included in the code. We are using all the 7 keys in the project by hashing keys in separate Class. Each time code runs into “**Too many requests**”, the code changes the API keys to next set of keys and resumes the code.

Step 4 : gender-api.com : function classify_gender_load_mongodb()

We have used an API called gender-api provided by gender-api.com. This API takes names and classifies the User gender. This API requires a key to connect to the server. The key is already included in the code in the function, under variable **myKey**

Extra key : ‘**BB8GI8Nu3dPWzG2NNVLPSeKo5hsFIcabCo5T**’

Step 5 : Loading Data into Mongodb

After preprocessing the data, we are loading the data to Mongodb , as an intermediate step so that we don't need to hit the API again. The function **classify_gender_load_mongodb()** classifies the gender loads the data into Mongodb. The connection string to the database is already provided in the same function.

Note : As we already loaded the data into Mongodb , we are not running the above again in our final code. But we have kept the function for the reference.

Step 6 : Run the code.

Please clone the entire project using the link :

cd <desired_location>

git clone <https://github.com/AkshayHebbar/SMDMFinalProject.git>

Once cloned , please run the file - **tweetrimony.py**

The End

