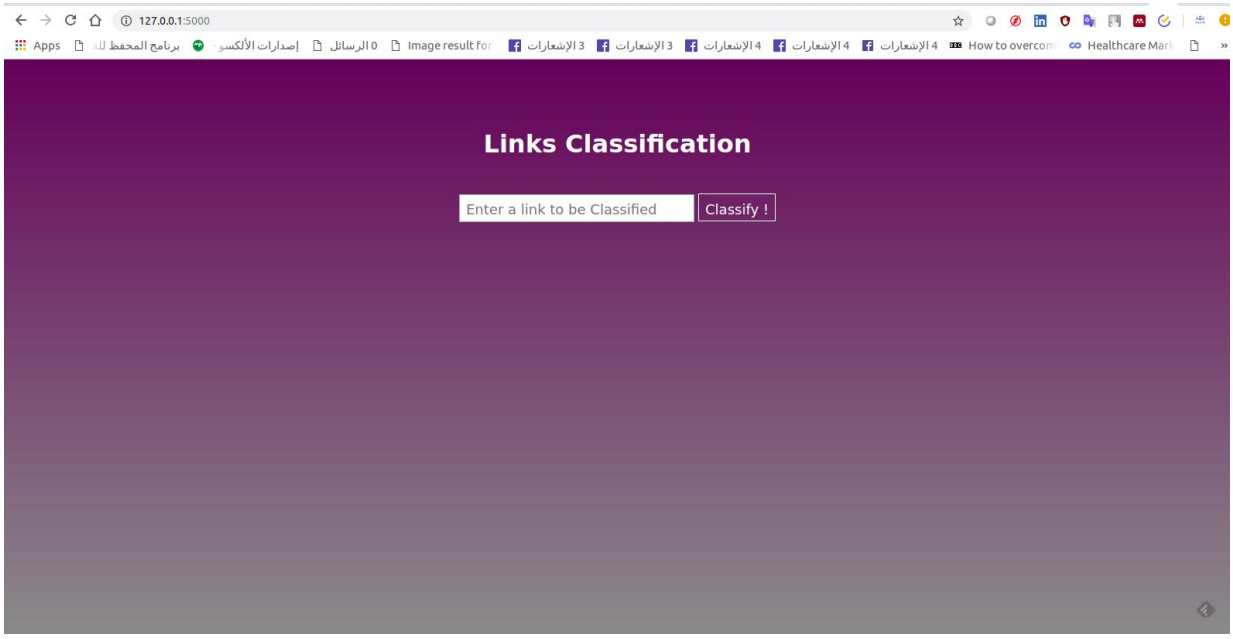
Links Classification

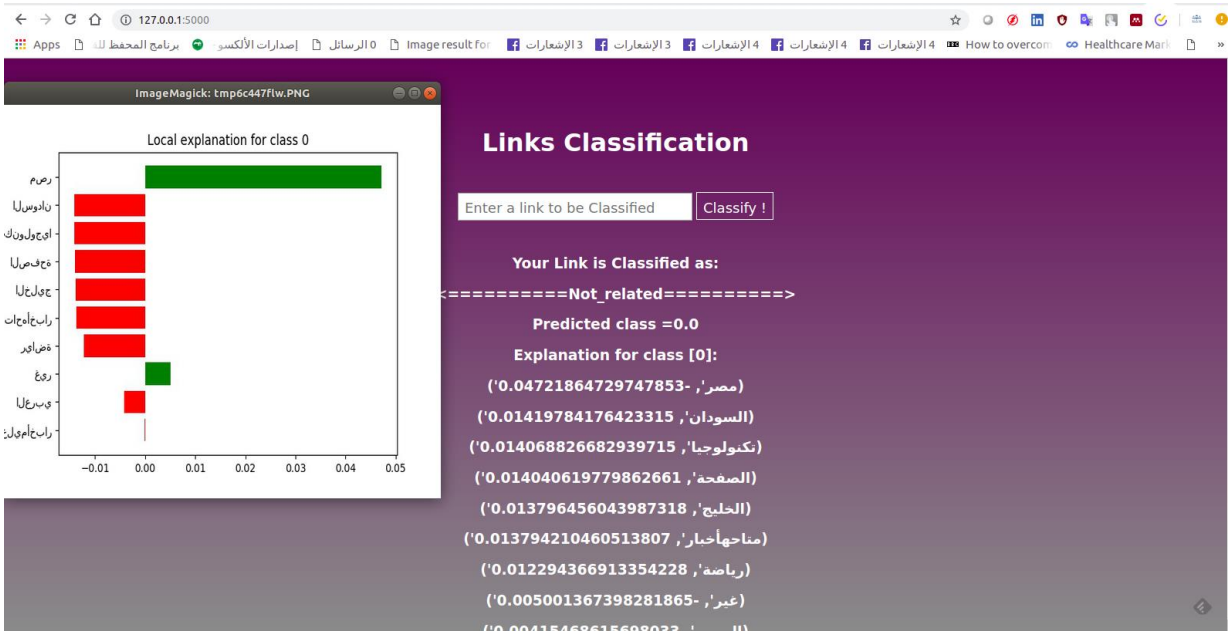
1. Overview:
    - The aim of this project is to classify links dataset to know if it is related to a particular TV series or not.
    - Dataset consist of two columns link and class where 1 means
    - related, and 0 means Not related.
    - Some of the links included in the dataset are not working which needs consideration while the feature extraction process.
    - Simple web application to run the script should be implemented.

2. Feature Engineering and Data Preprocessing:
    - Extracting title, keywords, and description for each link as features.
    - Solving the problem of the “TooManyRedirects” error by setting allow_redirects=False.
    - Solving the problem of not working links using try catch to catch the ConnectionError.
    - Cleaning the new dataset which contain the generated features.

3. Building Classification Model:
    - Using logistic regression to classify the links because:
        - I worked on a text classification project before and logistic regression achieved the best accuracy that’s why I decided to use it in this project. But if I had enough time, I would have tried more algorithms.
    - Building generalized model which takes the link as an input then generate the output which is "related" or "Not related" to the TV series.

4. Building Web Application:
    - Using flask to build a simple we application where you can provide your link and click classify to see the results in the same web page.





5. Resources:

https://medium.com/python-pandemonium/better-python-dependency-and-package-management-b5d8ea29dff1

https://stackoverflow.com/questions/15421746/matplotlib-writing-right-to-left-text-hebrew-arabic-etc

https://stackoverflow.com/questions/12570859/how-to-show-pil-images-on-the-screen

https://stackoverflow.com/questions/39772670/flask-return-multiple-variables

https://www.quora.com/How-do-I-read-an-image-in-Python

https://github.com/marcotcr/lime

https://stackoverflow.com/questions/21407147/python-requests-exception-type-connectionerror-try-except-does-not-work

https://towardsdatascience.com/machine-learning-nlp-text-classification-using-scikit-learn-python-and-nltk-c52b92a7c73a

https://stackoverflow.com/questions/12244057/any-way-to-add-a-new-line-from-a-string-with-the-n-character-in-flask?fbclid=IwAR2XwTIUuu79xkwH-WI_tz35yn3ThpqRwS8OHK_jMyhw5L6jkyWLRooNx44