

Links Classification

BY: ASMAA ALI ABDELWAHAB

Overview

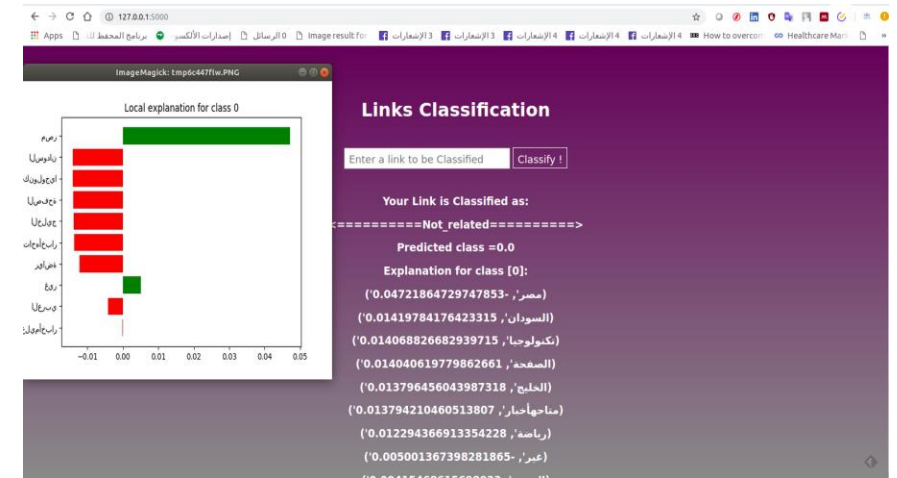
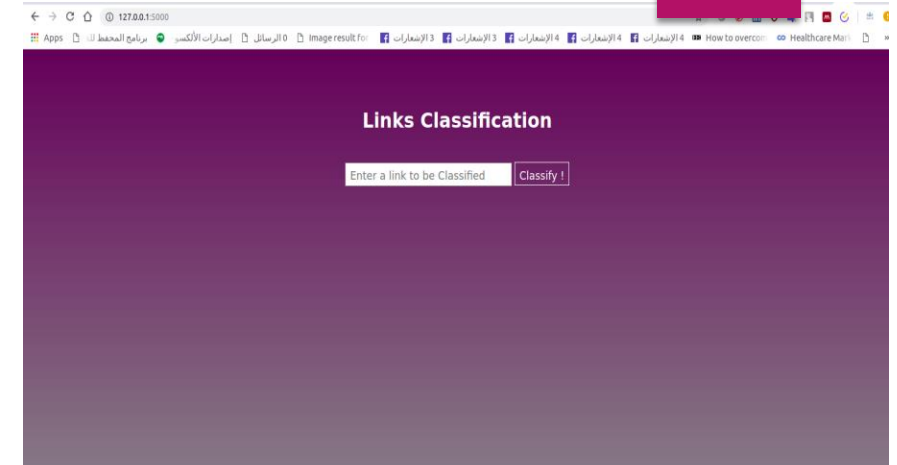
- ▶ The aim of this project is to classify links dataset to know if it is related to a particular TV series or not.
- ▶ Dataset consist of two columns link and class where 1 means related, and 0 means Not related.
- ▶ Some of the links included in the dataset are not working which needs consideration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

Approach

- ▶ 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.
- ▶ Using logistic regression to classify the links.
- ▶ Building more generalized model which takes the link as an input then generate the output which is "related" or "Not related" to the TV series.

Approach

- ▶ Using flask to build a simple web application where you can provide your link and click classify to see the results in the same web page.



Conclusion

- ▶ Classifying links according to the extracted features with accuracy above 85%.
- ▶ Building Web application which is easy to be used for classifying new links and displaying the classification details like scores given to the words by the classification model and a simple visualization of this scores.

Thank You
