NOVMBER 2021

Reddit Posts Topic Modelling (NLP)

Prepared by:

Abdultawwab Safarji
Najla Bin-Melha
Khalid Alrashed

Prepared for:

T5

INTRODCATION

This project works for Reddit to take posts in Reddit and do NLP topic modeling for each subreddit.

QUESTIONS

- To do NLP is a critical task to do if want machine to do this job we must ask important questions
- ·What are the most subreddit the affect specification?
- ·Does the model perform well on other samples?
- ·What are the best posts in the topic modeling?
- Why Riddet Toping Modeling?

SCOPE

- To Analyse Reddit the following information from previous samples were used
- Taken two subreddit randomly (Game& sport).

OBJECTIVE

- Reddit wants to divide posts in best topic modeling.
- To organize, understand and summarize large collections of textual information.
- Make it easy for the Reddit community to use unsupervised topic modeling and advance in technology as people discuss in text format.

TOOLS

Data manipulation: NumPy and Pandas
Modeling: Sklearn
Visualization: matlibplot and seaborn
NLP: PyCaret NLP , gensim, nltk,scipy
Others:
pyLDAvis, plotly, pylab,pickle, and regex