NOVMBER 2021

Reddit post topic modelling

Prepared by:

Abdultawwab Safarji
Najla Bin-Melha
Khalid Alrashed

Prepared for:

T5

INTRODCATION

We are a company that works for Reddit to take posts in Reddit and do the topic modelling for each subreddit.

QUESTIONS

To do NLP is critical task to do, if want machine to do this job we must ask important questions
·What are the most sub reddit the affect specification?
·Does the model perform well on other samples?
·What is the best posts in the topic modling?

SCOPE

To Analyse reddit the following information from previous samples were used
1- we toke two sub reddit randomly (Game& sport).

OBJECTIVE

- Reddit want to divide posts in best topic modelling.

TOOLS

Data manipulation: NumPy and Pandas
Modeling: Sklearn
Visualization: matlibplot and seaborn
Others:
NullFormatter, Plotly, Missingno, Filterwarnings, Counter. Orange
Dash, R Shiny.