Influencer analysis

# **Objective :** Find influences in twitter in Recent Trends

**PseudoCode/Procedure/flow/ Algorithm**:

**Step 1** : **This pipeline fetches data from twitter based on keywords. These keywords are fetched from create WordCloud which contains relevant keywords regarding the politics and sports.**

**WordClouds**: “”””””See References Section below””””””. Step1 references.

**Step 2**: **Cleaning process to remove noise from data after apply cleaning function. This function contains Stopwords, punctuation, emoticons and special characters removal from the data and will return clean tweet text.**

function cleaning(tweet\_text):

Remove stopwords , punctuation, stemming , lemmatization emoticons, special characters from the data.

return cleaned\_tweet

End

**Step 3**:

**Filtration process to filter irrelevant tweets from the data. To filter tweets data keywords and relevant keywords are fetched from the wordcloud and generate a filtration score for every processed tweet. This filtration score is generated based on no. of matched keywords in tweets.**

**Threshold for filtration process is greater than 1.**

Reason for threshold**: because at least one keyword should be there in tweet to prove the relevance of the tweet regarding the topics.**

function filtration(tweet\_text, keyword):

fetch filtrations data from data transform object model WordCloud

to calculate matching score and filter out tweets those

are below threshold of matching score required to get influencer users.

if tweet\_score > 1:

drop tweet from pipeline

else:

add tweet in pipeline

return survived\_tweets

End

**Step 4**:

**Sentiment Analysis to find the polarity of tweets that can be negative, positive and neutral. This sentiment analysis is achieved using text classification by applying various algorithms like naive bayes, SVC and logistic regression etc. In these algorithms naive bayes almost has better accuracy and performance, so naive bayes is included in pipeline for getting polarity for sentiment analysis.**

**This sentiment analysis function takes a clean text of tweets and process using selected naive bayes algorithm to provide sentiment, polarity score and subjectivity from text.**

function sentiment(tweet\_text):

Find sentiment from the tweet text and categorize into either negative tweet or positive tweet.

Also calculate polarity and subjectivity of tweet to update in pipeline.

return sentiment, polarity, subjectivity

End

**Step 5**:

**To get influencer out of processed tweets some profiling is done using favorites counts, retweet counts and sentiment for profiling. After getting cleaned and preprocessed data threshold values are extracted for favorites counts and retweet counts.**

**Threshold values are extracted by getting average from favorites and retweet counts. Above average values are set to threshold values.**

**Because above average counts that tweet can be categorized into some kind of influencer category.**

**This function takes tweets and its info for processing and return only survived tweeets which can be considered as influencer and can be categorized into negative influencer or positive influencer.**

function findInfluencial\_users(tweet):

Find most influenctial users based on different criterias like most

favorites counts, retweet counts, influence type, profiling etc.

return most influencial users

End

**References**

**Step1: WordClouds for politics and sports**

**"politics": ["government", "campaign", "delegate", "expedition", "incumbent", "politics", "indian", "bjp",**

**"congress", "india", "political", "party", "poll", "suicide", "majority", "vote", "election",**

**"parliament", "constituency", "candidates", "constituent", "prime", "minister", "cabinet",**

**"mp", "manifesto", "affairs", "deliberative", "liberal", "democratic", "politicize", "office",**

**"failure", "modi", "gandhi"],**

**"sports": ["athletics", "sport", "competition", "indian", "india", "game", "racing", "gymnastics", "soccer",**

**"football", "sportsman", "offside", "cycling", "tennis", "cricket", "captain", "bcci", "icc",**

**"run", "team", "archery", "baseball", "frisk", "coach", "champion", "chess", "english", "field",**

**"gameday", "olympics", "snowboard", "league", "plan", "stadium", "world", "cup", "playground",**

**"hockey", "pitch", "court", "fitness", "venue", "event", "employment"]**

**//word cloud is prepared from which data because “english” is keyword in sports section and this is not relevant.**

**Step2:**

[**https://medium.com/@dobko\_m/nlp-text-data-cleaning-and-preprocessing-ea3ffe0406c1**](https://medium.com/@dobko_m/nlp-text-data-cleaning-and-preprocessing-ea3ffe0406c1)

[**https://towardsdatascience.com/nlp-for-beginners-cleaning-preprocessing-text-data-ae8e306bef0f**](https://towardsdatascience.com/nlp-for-beginners-cleaning-preprocessing-text-data-ae8e306bef0f)

[**https://towardsdatascience.com/effectively-pre-processing-the-text-data-part-1-text-cleaning-9ecae119cb3e**](https://towardsdatascience.com/effectively-pre-processing-the-text-data-part-1-text-cleaning-9ecae119cb3e)

**Step3:**

**Proposed process for filtration**

**Step4:**

[**https://monkeylearn.com/sentiment-analysis/**](https://monkeylearn.com/sentiment-analysis/)

[**https://www.lexalytics.com/technology/sentiment-analysis**](https://www.lexalytics.com/technology/sentiment-analysis)

**Step5:**

**Proposed process to find influencer and profiling**

**Code explanation and setup**

**Step1:** Download code repo from github

**Step2**: Create virtual environment for project.

**Step3:** install requirements from /main/resources folder by using command:

pip install -r requirement.txt

**Step4**: Run server.py

“””””””””””””””””””””output””””””””””””””””””””””””””””

**Raw folde**r in directory: /main/resources

Contains Raw files

**Data folder:** /main/resources

Contains preprocessed data after cleaning process and filtration process.

**Reports folde**r: /main/resources

Contains final influencer data with profiling.

Model\_info.json will contains models accuracies for comparisons.

**Training\_data folder /main/resources**

**\*\*\*\*\*Changes these above paths in config.env file in the resources folder as per your system directory\*\*\*\*\***