

CS432/532: Final Project Report

Project Title: Analysis of GoT's Twitter Dataset

Team Member(s): Akarsh Kancharla, Keerthan Vudutha

I. PROBLEM

Game of thrones is one of the biggest television series in world which ended in 2019. Unexpectedly the cast of the game of thrones expressed their mixed opinions on the final season even before the premier, considering that the previous seasons were well received by audience it is actually unbelievable to say that final season is not well received by many, we are trying to analyze about final season on a social media platform

Social Media Platforms have become a very good medium to know how the receiving end behaves in response to your products or services. People speak about things on social media fearlessly. So with the help of data from twitter and after analyzing that data one can tell a lot about how people received the final season.

This project uses pre collected twitter data set of GoT's final season. In the first half of the project we find tweets that are recently tweeted by the users in the dataset, frequency of negative or positive words that are mentioned in tweets about game of thrones and popular characters in the show.

The second half is about sentiment analysis on the dataset i.e finding episode wise reaction of audience, nature of the distribution of dataset and comparing the twitter users reactions and critics reviews. we retrieve data from database using MongoDB and various functions from python libraries are used for analysis.

II. SOFTWARE DESIGN AND IMPLEMENTATION

A. Software Design and NoSQL-Database and Tools Used

We took a twitter dataset related to game of thrones season 8, The dataset is in CSV format, this is truncated and cleaned into a new CSV file that is suitable for project implementation. This new CSV file is inserted as a collection into MongoDB. Reviews of critics from Rotten Tomatoes are collected into a separate CSV file. This CSV file is inserted as new collection into the same database of MongoDB. A Bag of positive and negative words are collected from another source and used for the purpose of analysis. Some python libraries are also used for analysis in this project.

Tools Used:

Database: MongoDB

Programming Language:Python

Used Python Libraries: *Pymongo, nltk, textblob, datetime, matplotlib, scipystats, numpy, array*

B. Supported Queries

1. Most recent tweets in dataset:

Returns text of the top 5 tweets that have latest time stamps

2. Most frequently used positive word:

Returns top 5 frequently used positive words with number of times they got mentioned in whole dataset

3. Most frequently used negative words:

Returns top 5 frequently used negative words with number of times they got mentioned in whole dataset

4. Most frequently mentioned character:

Returns top 5 frequently mentioned characters with number of times they got mentioned in whole dataset

5. Find and plot sentiment analysis on weekly basis :

Returns sentiment scores of tweets by weekly basis and visualize them.

6. Find skewness of the sentiment scores:

Returns the skewness value for the above calculated sentiment scores

7. Find Standard Deviation of sentiment scores:

Returns the value of standard deviation in sentiment scores for whole data set and also divides the sentiment scores of dataset by week basis and plots weekly standard deviation.

8. Finding relevance of twitter data with other sites reviews using correlation:

Return correlation value between sentiment scores of twitter dataset and sentiment scores of critics reviews

9 CRUD operations:

- Able to insert new tweets into database
- Able to delete the most recent tweets in database
- Able to edit the recent tweet of a specific user in the database

III. PROJECT OUTCOMES

https://github.com/keerthanvs/Db_project3

<https://www.youtube.com/watch?v=bnY4JFRJEPU&t=503s>

REFERENCES

- [1] The Dataset is obtained from the following source on the internet
<https://www.kaggle.com/monogenea/game-of-thrones-twitter>.
- [2] Bag of positive, negative words are taken from
<https://github.com/jeffreybreen/twitter-sentiment-analysis-tutorial-201107/tree/master/data/opinion-lexicon-English>
- [3] Critics reviews for Game of Thrones final season are collected from
https://www.rottentomatoes.com/tv/game_of_thrones/s08
- [4] <https://api.mongodb.com/python/current/tutorial.html>