

Online Show Popularity Prediction

Triveni Landge ^{#1}, Supriya kore ^{#2}, Neha Bora ^{#3}, Priya Shelke^{#4}

Information Technology Department, Vishwakarma Institute of Information Technology

Abstract

Reality show is the new mantra of television producers and channel executives. Online Show Popularity prediction using sentiment analysis is one of the most interesting tasks because of the “similar -but-tweaked-here-and-there” shows churned out by the competition. Nowadays number of productions and cost of production is very high due to which it become most important and challenging tasks to predict Online Show Popularity using some simple method over traditional one. These traditional methods are costly and consume time as well. Online Show producers and advertisers constantly seek ways to understand their target audience along various dimensions. This allows Online Show producers to understand certain aspects of the preferences of their viewers, gain valuable feedback and improve their shows. The accurate and timely prediction of program popularity is of great value for content providers, advertisers and broadcast operators. This information can be beneficial to evaluate the performance of the Online Shows and also calculate how many people like a particular show or actors of that show and predict popularity of those shows, based on the text reviews.

Keywords: *Tweets, Popularity prediction, Sentiment Analysis, Online Shows, Natural Language Processing, Porter Stemmer.*

I. OVERVIEW

Presently many of the online shows are reality shows which particularize in dancing, singing and acting and so on. An attempt is made to build such a system that will interpret people's sentimental comments based on Online Shows. The comments that are tweets associated to the specific show will be extracted as an input. The comments will be gathered from social networking website named Twitter. On the basis of tweets given by the people, the Show popularity will be predicted. Based on the people's comments and ratings, visitors can view Online Show popularity data. Visitor can also view the rating of a popular show as well as view the top shows.

II. INTRODUCTION

Television industries in India are estimated to be worth billions of dollars. Analysis of the social media is becoming a very important instrument to monitor communities, user's preferences and to make predictions. Among the social media solutions, Twitter is one of the most widespread micro blogs allowing users to have a personal news feed and followers attached to it.

There are so many different types of genres for our Online Shows. Different people like to watch different genres and always have a preference. A show's popularity also depends on what genre it belongs to. For example, a thriller might be popular among the younger crowd while a drama will fail for that same crowd. The popularity analysis of these Online Shows will help both the audience and the corporate advertisers. Audience will be able to make choices based on the suggestions about which show to watch next and the corporate advertisers will be able to sell ads based on the popularity of these shows.

The purpose of this system is to evaluate the performance analysis of Online Show and also calculate how many people like a particular show or actors of that show and make popularity prediction of that particular show, based on the text reviews.

The paper is organized as follows section 1 and 2 describes the general ‘overview’ and ‘introduction’ of the performance analysis. Section 3 holds the ‘specific objective’ to design a system that determines Online Show Popularity based on sentiment analysis of people on Twitter. Section 4 ‘literature survey’ provides a description of the methods that are used for the Online Show Popularity Prediction. ‘Implementation details’ are elaborated in section 5 with the ‘proposed system’ and ‘algorithm used’ in the proposed system. This section also contains the proposed ‘system architecture’ and the example of the algorithm used in the proposed system. ‘Conclusions’ are drawn in Section 6.

III. SPECIFIC OBJECTIVES

- To investigate the problems associated with the current methods used by television industry in gauging the popularity of a show.
- To evaluate algorithms used in data mining and prediction of Online Show’s popularity.
- To design a model that determines a show’s popularity based on sentiment analysis on Twitter.
- To test the Online Show Popularity predictive model.

IV. LITERATURE SURVEY

In this section, we have presented brief review on points stated below:

Motion picture Success Prediction utilizing Data Mining

In 2017, Javaria Ahmad, Amr Yousef and Bill Buckles [4] applied data mining processes to extract trends and patterns which can be beneficial for movie success prediction. Information mining manages examples of given information and finding patterns. It recognizes the concealed examples and connections among different factors.

Big Data Analytics

Egebjerg, Niklas Hedegaard, Gerda Kuum and Raghava Rao Mukkamala [5] in 2017 used Big Data Analytics to anticipate fan commitment as far as onlookers and television appraisals. It was workable for them to set up a prescient model for the quantity of onlookers and audience members decently depending on just two control factors to be specific match type and match day. The work had impediments including the way that not many matches were played amid the example time frame, no refinement was made between positive/impartial/negative posts and the social information was accessible for a long time.

The paper [7] describes the results of a research work aimed at identifying suitable predictive models to predict media show audience (number of people following the programmer) by exploiting social media info for reality shows. The research signifies the analysis of their validity in terms of prediction performance. The rating can be affected by the presence of other competing TV programs in the same time slots of social media analytics for network television ratings.

Social Media Analytics for Network Television Ratings

In this study [2], social media data for 38 programs was collected for a period of five weeks. Here, what was extracted was the number of tweets, followings, followers, likes and mentions of each program (Oh et al., 2015). The key purpose of the assignment was to identify the relationship between posts by the official account of a program and its rating performance. Here, an analytics framework was proposed, which entails collecting data, cleaning data, extracting features and carrying out analysis.

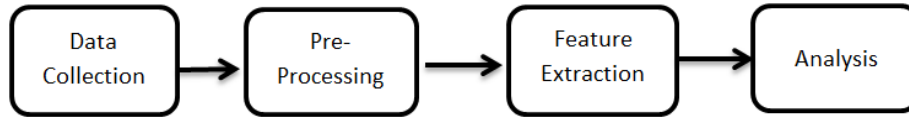


Fig. 1 Proposed Framework for Social Media Analytics (Oh et al., 2015)

In [6], a very high-level analysis of the twitter data related to TV programmer has been proposed, showing that the degree of interaction on Twitter was correlated with X Factor programmer and its evolution. The approach of using Twitter for TV programmer analysis is also used by Nielsen for analyzing if Twitter is helping the audience or vice versa.

In [8], averaged value of audience in past events and Twitter data (contributions per minute) have been used for predicting audience on successive political shows having long series of events; therefore, demonstrating a correlation among the volume of tweets and the audience.

In [5], a solution to predict football game results has been proposed by considering the volume of tweets. Hence the achieved prediction rate was in the range of 68%.

Sentiment analysis for social media data

This study [3], (Ramadhani & Goo, 2017) intended to analyze the data. Here, the first step is text mining, whereby 3 steps are involved:

- i. Retrieving information from social media.
- ii. Performing partial analysis and identifying any relevant data points.
- iii. Digging into the data to retrieve insights.

A dataset of about 4,000 tweets was used on a neural network which entailed a hundred neurons, 3-layer architecture as well as a stochastic descent. The layers provide a step-by-step method of clearing words, sentences and finally dictating the popularity.

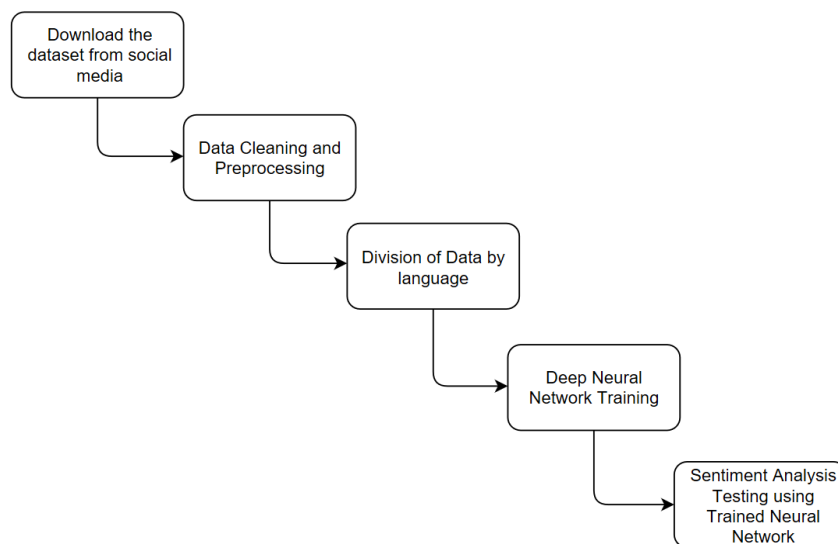


Fig. 2 Deep Learning on Data from Twitter

Prediction of Movie success using Twitter Sentiment Analysis

In this study [1], the popularity of movies is determined by carrying out sentiment analysis on tweets related to the movie title. Eight movies were predicted. The metrics used were hit, flop or average. Since the numbers of tweets used to perform the study were many thus required lots of labour to manually label them, 200 tweets were selected on a random basis (Jain, 2013).

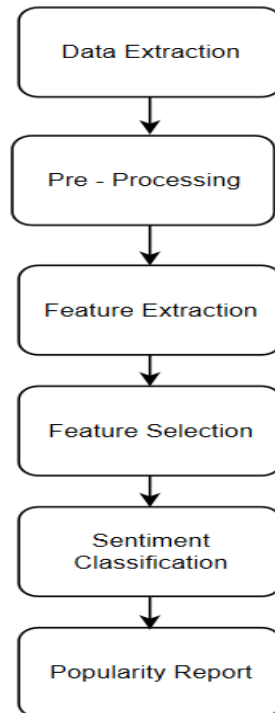


Fig. 3 The process of sentiment analysis

In addition to this, a threshold was set. For a movie to qualify as a hit, its PT-NT ratio needed to be more than or equal to 5. On the other hand, for it to qualify as average, the same ratio needed to be less than five but more than 1.5, whereas less than 1.5 was considered a flop (Jain, 2013). The prediction of the popularity was done by first training the Naïve Bayes classifier through the use of the NLTK toolkit, while the feature used was word count.

V. IMPLEMENTATION DETAILS

A. PROPOSED SYSTEM

Figure for the architecture of the Proposed System is given below. The system flow is divided into followings way. Initially the system fetches Online Show related data from Twitter API where text reviews are related to the particular show and actors of that show.

1. Get Online Show Data from Twitter:

- In order to create the hash tagged data set, we first filter out certain tweets such as duplicate tweets, non-English tweets and tweets that do not contain hashtags.
- From the remaining set, we will investigate the distribution of hashtags. After this, we identify what we desire will be sets of frequent hashtags which indicate positive, negative and neutral messages. These hashtags are used to select the tweets that will be used for development and training purpose further.

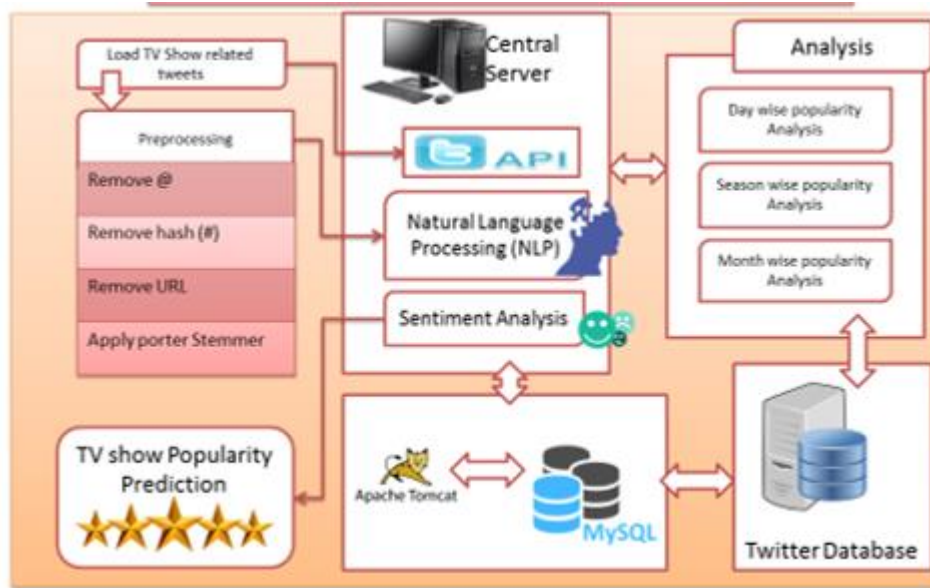


Fig. 4 Proposed System Architecture

2. *Preprocessing:*

i. Tokenization

It splits sentences into words. The process of breaking a stream of text or sentence up into words, phrases, symbols or other meaningful elements called tokens and this entire process is known as tokenization. The input is nothing but the list of the tokens for further processing such as parsing or text mining.

ii. Normalization

- a. Removes stopwords from input text data.
- b. Stopwords are words that are automatically omitted from a computer-generated concordance or index.

iii. Part-of-speech (POS) tagging

- a. Detects if the word token is noun, verb, adjective
- b. POS Tagging in which a word is assigned in accordance with its syntactic functions. In English the main parts of speech are noun, pronoun, adjective, verb, preposition, adverb, conjunction, and interjection.

3. *Apply NLP:*

Some of the main challenges that are included in natural language processing are teaching computers to understand the way humans learn and use language. Some search engines including Google base its machine translation technology on NLP deep learning models. This permits algorithms to read text on a webpage, interpret its meaning and translate it to another language.

4. *Sentiment Analysis:*

The term Sentiment analysis is another primary use case for NLP. By applying sentiment analysis, data scientists can assess comments on social media to see how their business's brand needs to perform or the current performance, for example, review notes from customer service teams to identify sector where people want the business to perform better or for brand advertisement.

5. *Analysis:*

The Online Show Popularity Analysis is done in following 3 ways:

Table 1
 Online show popularity analysis in 3 different cases

Day Wise Popularity Analysis	Season wise Popularity Analysis	Month wise Popularity Analysis
The analysis for one day show is where a user might want to find particular date wise analysis.	This analysis includes in three seasons which type of shows was to hit in given particular period this type analysis done in season wise analysis.	It is also possible to calculate month wise popular shows list and watch it. Like day and season wise analysis we analyze a month wise popularity of a show.

B. *ALGORITHM USED:* -

PORTER STEMMER ALGORITHM

The stemming algorithms work by trimming the end or the beginning of the word, taking into consideration a list of common prefixes and suffixes that can be found in an reflected word. Following are the steps of this algorithm: -

Steps of Algorithm:

- Step1. Trim the plurals and -ed or -ing suffixes
- Step2. Converts terminal y to i when there is another vowel in the stem
- Step3 Mapping of double suffixes to single suffix: -ization, -ational, etc.
- Step4. Deals with suffixes, -full, -ness etc.
- Step5. Takes off -ant, -ence, etc. Removes a final -e

Example of Algorithm:

Table 2
 Example of Porter Stemmer Algorithm

Form	Suffix	Stem
Studies	-es	Studi
Study ing	-ing	Study

VI. RESULTS

Following table and graph represents the Algorithm comparison with respect to its accuracy.

Table 3
 Algorithm Comparison

Algorithm wise Accuracy Comparison				
Sr	Algorithms	No. of Reviews	Accurate	% Accuracy

No.			Answer	
1	NLP Algorithm	150	142	0.946666667
2	Naïve Bayes	150	121	0.806666667
3	KNN	150	133	0.886666667

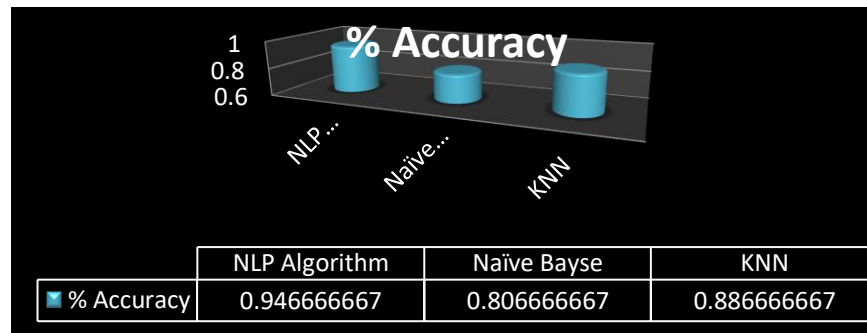


Fig. 5 Algorithm wise Accuracy Comparison.

VII. CONCLUSION

The research is purposed to design a system that determines the popularity of Online Show based on sentiment analysis on Twitter data. To be able to successfully deliver on the set specific objectives, intense research on relevant works and literature survey was done. Research on the available algorithms was done to clearly understand how they can be applicable in the research. Tweets related to Online Shows are collected based on a set of predetermined Twitter hash tags using the Twitter API, and then text pre-processing is applied on the data. The paper also illustrates that sentiment analysis can be useful when intending to get data on the popularity of Online Show using the available related tweets on Twitter. The popularity analysis is done on three aspects that includes: Day Wise, Season wise and Month wise Popularity Analysis to get accurate results.

REFERENCES

- [1] Jain, V. (2013). Prediction of movie success using sentiment analysis of tweets. *The International Journal of Soft Computing and Software Engineering*, 3(3), 308-313.
- [2] Oh, C., Yergeau, S., Woo, Y., Wurtsmith, B., & Vaughn, S. (2015, April). Is Twitter psychic? Social media analytics and television ratings. In *2015 Second International Conference on Computing Technology and Information Management (ICCTIM)* (pp. 150-155). IEEE.
- [3] Ramadhani, A. M., & Goo, H. S. (2017, August). Twitter sentiment analysis using deep learning methods. In *2017 7th International Annual Engineering Seminar (InAES)* (pp. 1-4). IEEE.
- [4] Ahmad, J., Duraisamy, P., Yousef, A., & Buckles, B. (2017, July). Movie success prediction using data mining. In *2017 8th International Conference on Computing, Communication and Networking Technologies (ICCCNT)* (pp. 1-4). IEEE.
- [5] Egebjerg, N. H., Hedegaard, N., Kuum, G., Mukkamala, R. R., & Vatrapu, R. (2017, June). Big Social Data Analytics in Football: Predicting Spectators and TV Ratings from Facebook Data. In *2017 IEEE International Congress on Big Data (BigData Congress)* (pp. 81-88). IEEE.
- [6] Lochrie, M., & Coulton, P. (2012, January). Tweeting with the telly on!. In *2012 IEEE Consumer Communications and Networking Conference (CCNC)* (pp. 729-731). IEEE.

- [7] Hsieh, W. T., Chou, S. C. T., Cheng, Y. H., & Wu, C. M. (2013, October). Predicting tv audience rating with social media. In *Proceedings of the IJCNLP 2013 Workshop on Natural Language Processing for Social Media (SocialNLP)* (pp. 1-5).
- [8] Giglietto, F. (2013). Exploring correlations between TV viewership and Twitter conversations in Italian political talk shows. *Available at SSRN 2306512*.