

# Statistical Analysis on Movie Reviews and Ratings

<sup>1</sup>Veeresh Belgur, <sup>2</sup>Aniket Karande, <sup>3</sup>Nikhil Kulkarni, <sup>4</sup>Pranil Nalawade <sup>5</sup>Prof. Aniket M. Junghare

<sup>1,2,3,4</sup>BE in Computer Engineering ZCOER, Pune, Maharashtra, India

<sup>5</sup>Professor at ZCOER, Pune, Maharashtra, India

**Abstract**— In this paper, we describe how Statistical Analysis on Movie Reviews and Ratings can be done. PEOPLE's opinion is one of the most important sources for different services. In particular, Social-Media response have turned into a kind of virtual currency for businesses looking to market their products, identify new opportunities, and manage their reputations. On the other hand Statistics representation gives the clear idea about the products or things. Statistical Analysis on Movie Reviews and Ratings gives users a perfect picture what social media thinks about the movie. A complete overview is presented here. The movie-rating information which are going to generate is based on various sources like, Twitter, Facebook, IMDb and Google Trend.

**Index Terms**— Natural Language Processing (NLP), Text Mining, Sentiment Analysis, RHadoop.

## I. INTRODUCTION

Practically, when we are not familiar with a specific product, we ask our trusted sources to recommend one. Today, the popularity of the internet drives people to search for other people's opinions from the Internet before purchasing a product or seeing a movie. Many websites provide user rating and commenting services, and these reviews could reflect user's opinions about a product or any particular movie. For example, Amazon.com list out the number of reviews, the percentage for different ratings, and comments from reviewers. When people want to purchase a book, CDs, or DVDs, these comments and ratings usually influence their purchasing behavior.

In addition to these websites, a social media networking is another important source for people to search for other people's opinions like user can have the review of movies through Tweets on Twitter and Facebook comments or check in. The user can have the review from the critics of various different movies and poll voting.

Essentially, the task of determining whether a movie review is positive or negative is similar to the traditional binary-classification problem. Given a review, the classifier tries to classify the review into Positive or Negative category. However, opinions is a natural language are usually expressed in subtitle or in complex way.

In this paper we are going to collect movie reviews from Social media i.e. twitter and facebook that do not consist of any rating

information. Sentiment analysis is performed to determine the semantic analysis of the reviews and movie-rating score is based on the sentiment-analysis result.

Summarization technique is used to reduce the size of information. The system will summarize the reviews i.e. positive reviews and negative reviews and provide the user an overview for all the reviews. Through this user will decide which movie he/she wants to look. All the result is presented in single page which will helpful for the user to understand better and easily.

## II. EXISTING SYSTEM

Currently available Rating system in market are mainly based on three things, i.e. Critics, Opinion poll & Marketing price(Money issued). Popular websites like Bookmyshow, IMDB provides the brief details but not in statistical format. They only provide likes/dislikes which are given on their particular website.

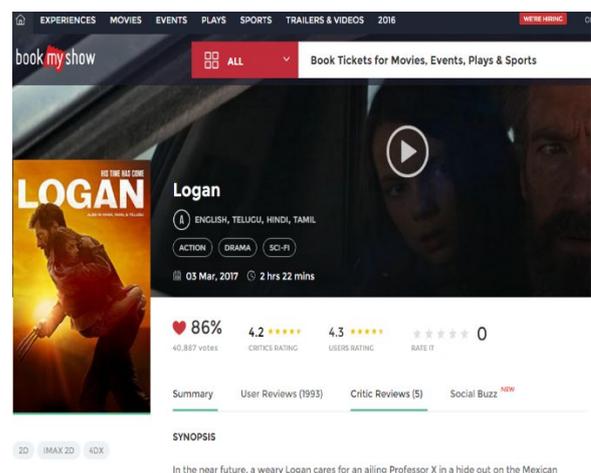


Fig 1: Bookmyshow rating view

<https://in.bookmyshow.com/buytickets/logan-/movie--ET00047340-MT/20170401>



Fig 2: IMDB rating view

<sup>2</sup> [http://www.imdb.com/title/tt3315342/?ref\\_=fn\\_al\\_tt\\_1](http://www.imdb.com/title/tt3315342/?ref_=fn_al_tt_1)

Here we are proposing totally new way of ratings which will be very useful for user/movie lovers from which they can get detailed and all in one detail about the movie. Currently not a single website is hosted which covers all the social media data into one single page. So these is completely concept.

### III. PREFERRED TECHNOLOGY

An enterprise weblog analysis system based on the reviews, which are fetched from various sources using the R Programming interface. Using the provided tokens of Facebook and Twitter, dynamic analysis of the reviews for particular movie is been done from Facebook and Twitter. And using Shiny web-app we are fetching the Youtube Trailer and Likes/Dis likes for a particular Movie. Using the particular API all this review, Youtube Trailer, Google Trends will be collaborated. By using Shiny web-app API we are displaying all the Reviews ,Trailers, Trend or impact of particular Movie in particular region.

Shiny is an open source R package that provides an powerful web framework, elegant and useful for building web applications using R. Shiny helps to turn your analysis into interactive web applications without requiring the existence of HTML, CSS, or JavaScript knowledge.

#### 3.1 Packages Used

*library(ggplot2)* : ggplot2 is used in plotting system for R, it is based on the grammar of graphics, which take the good parts of base.

*library(dplyr)* : dplyr is the next version of plyr, mainly focused on tools for working with data frames. It identify the most important data manipulation tools needed for data analysis and make them easy to use from R.

*library(tm)* : It is used for text mining purpose.

*library(twitteR)* : Provides an interface which used to get access the Twitter web API.

*library(Rfacebook)* :To connect with facebook and fetching data.

*library(shiny)* :Provides user interface like html to build webpage. Normally it consists two files ui.R & server.R.

*library(shinydashboard)* : It helps to produce result in dashboard format.

*library(gtrendsR)* : Google trends package to access data directly from google trends.

*library(wordcloud)* : Helps to produce twitter Wordcloud i.e most commonly used words in tweets.

### IV PROPOSED SYSTEM

As we are using Twitter and Facebook for review summarization it has some of the special featurization. Statistical Analysis on Movie Reviews and Ratings gives users a perfect picture what social media thinks about the movie. A complete overview is presented here.

We are going to provide wordcloud, sentiment analysis based on movie tweets (positive/negative/neutral) using graph and pie-chart in twitter. In facebook we mainly focus on the count of check-in for a particular movie and number of likes on that movie's official facebook page. If we talk about Youtube we are going to provide movie trailer as well as likes, dislikes of that particular movie trailer and number of views of that movie trailer. And about Google Trends we are going to show interest over time and interest over region that is what impact the movie has kept on particular region and for how much period. We are providing description about movie using the dataset, which will be fetched from IMDB using their API.

All the result is presented in single page which will helpful for the user to understand better and easily.

### V. DATA SOURCE

Data is collected the movie reviews from twitter, facebook. . Since the original data is in an tweet format we need tables to store the data ,We need to remove hash-tags from the tweet. Training data is necessary for manual classification is performed to classify the training reviews into positive or negative reviews. We selected trending tweets from the twitter as the data. In addition to the model-building data, we collected trending tweet movie from the tweeter, and these tweet will be used for rating of movie.

We are also including the dataset from IMDB for the description of movie. This description as access through APIs and tokens. This dataset is used for limited and basic information as like movie name, actor, runtime, released date, director, rated,etc.

## VI. CONCLUSION

So from the above analysis it is concluded that there is a need for statistical analysis of movie ratings and reviews. This would be great and unique concept that will introduced in market. Same concept can be extended to other products like mobile phones or any product to detrmine trends in social media.

## REFERENCES

- [1] Chien-Liang Liu and Wen-Hoar Hsaio, "Movie Rating and Review Summarization in Mobile Environment," in IEEE Transactions On Systems, Man, And Cybernetics—Part C: Applications And Reviews, Vol. 42, No. 3, May 2015
- [2] Walid Maalej and Hadeer Nabil, "Bug Report, Feature Request, or Simply Praise? On Automatically Classifying App Reviews," in IEEE RE 2015 - Authors' version of the work (University of Hamburg), 2015.
- [3] BOOKMYSHOW ("Bookmyshow"), "Bookmyshow". *BookMyShow*. N.p., 2017. Web. 31 Mar. 2017. WEBSITE
- [4] IMDB - MOVIES, TV AND CELEBRITIES, ("Imdb - Movies, TV And Celebrities"), "Imdb - Movies, TV And Celebrities". *IMDb*. N.p., 2017. Web. 31 Mar. 2017. WEBSITE
- [5] Jyoti, Dr. Sanjeev Dhawan, "Analysing User Ratings for Classifying Online Movie Data Using Various Classifiers to Generate Recommendations," in IEEE, 2015.
- [6] Maruthi Prithvirajan , Vivian Lai Analysis of star ratings in consumer reviews: A case study of Yelp, in IEEE, 2015.
- [7] Pooja D. Savant, Debnath Bhattacharyya, Tai-hoon Kim, "Hadoop based Weblog Analysis: A Review," in International Journal of Software Engineering and Its Applications 2016.

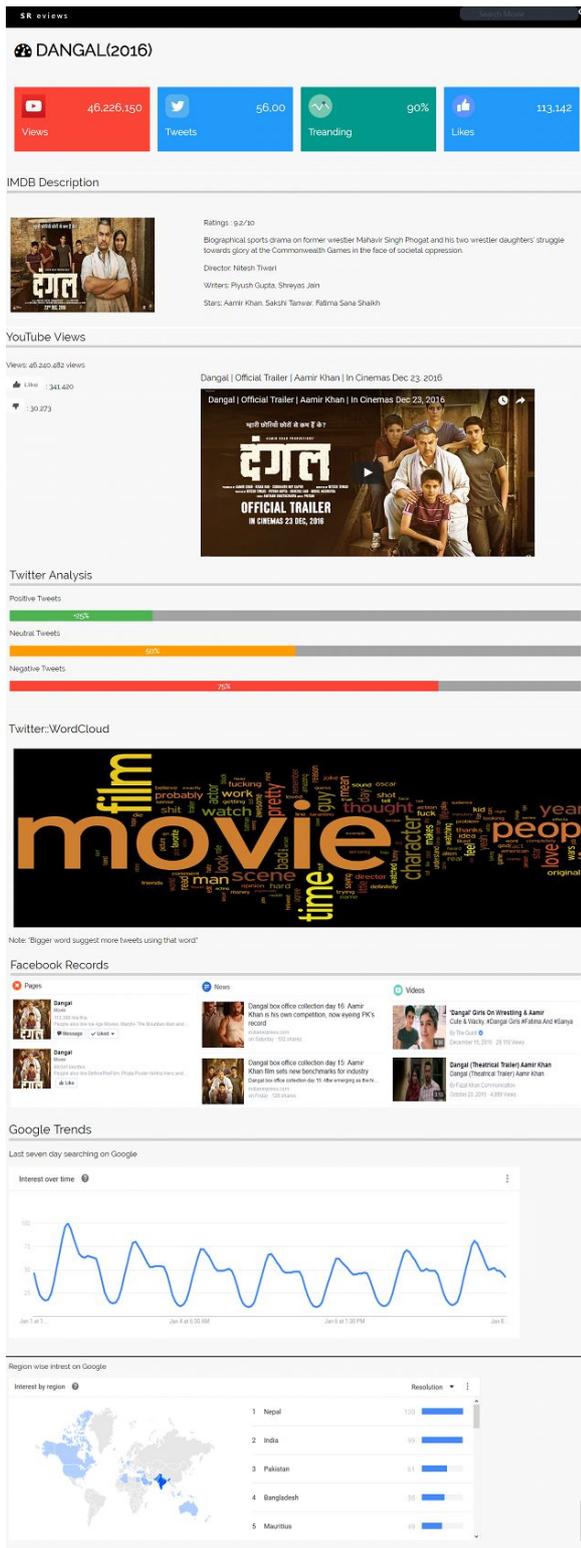


FIG 3: RESULTS OF STATISTICAL ANALYSIS