A Survey on Prediction of Movie's Box Office Collection Using Social Media

Vani Sahu*
PG Scholar, CSE, VITS, Bhopal, India

Prof. Sumit Sharma
HOD CSE, VITS, Bhopal, India

Abstract
Predicting the box office profits of a movie prior to its world wide release are a significant but also an exigent problem that needs a advanced of Intelligence. Currently, social media has given away its diagnostic strength in a variety of fields, which encourages us to develop social media substance to predict box office profits. The collection of movies in provisions of profit relies on so many features for instance its making studio, type, screenplay superiority, pre release endorsement etc, each of which are usually utilized to approximation their probable achievement at the box office. Nevertheless, the "Wisdom of Crowd" and social media have been accredited as a powerful indication in appreciative customer activities to media. In this survey, we converse the influence of socially created Meta data derived from the social multimedia websites and review the effect of social media on box office collection and success of movies. This survey paper is written for (social networking) investigators who looking for to evaluate prediction of movies using social media. It gives a complete study of social media analytics for social networking, wikis, actually easy syndication feeds, blogs, newsgroups, chat and news feeds etc.
Keywords: Social Networking, Social media. Movie's Box Office, Prediction, Profitability, Sentiment analysis

1. INTRODUCTION

Movie box-office profits specify the economic collection of movies. The capability to analyze a movie's box office profits before its theatrical release might diminish the producers' financial threat. Often, to amplify a movie's financial achievement, huge advertising deals are made to sponsor it before its theatrical release. Consistent predictions could notify production judgments in previous phases of the production procedure, with providing guidance for movie viewers. Although movie's success has been considered an unpredictable problem [1], several studies have attempted to develop approaches for movie box-office profits prediction [1, 2]. Specifically, the construction of computational models to investigate the relationships among movie-related variables has achieved considerable success. Pre-release movie-related data such as genres, ratings, and participating actors, are already extensively exploited for such predictions.

In recent times, consideration has been drawn to the potential for leveraging further movie related statistics to develop the prediction collection [3]. Specifically, movie posters include information that could concern movie box office profits. As an advertising standard, movie posters are purposely designed to express the content of movies and magnetize the concentration of probable viewers. Accordingly, movie posters are frequently recognized by viewers extensive before the movie's theatrical release. Gradually more, movie posters are made publicly accessible on online movie databases (e.g., IMDB), so it's significant to explore the influences of movie posters on movie box office profits. Movie posters have attracted research interest [4] in the field of movie genres classification. Since the content of movie posters is extremely related to their genres, the enclosure of features of movie posters with other movie related data might give more information for predicting movie box office profits. Nevertheless, feature withdrawal in these techniques is typically complete by hand, which doesn't simplify well and relies on previous facts.

The precise prediction of movie box office profits is extremely important for the drop of market threat, enhancement of the management of the film industry, and promotion of the development of a film-related derivative product market [5]. However, predicting movie box-office profits is a challenging problem, as it's very difficult to discover the essential reason for the volatility of the movie box-office profit [1]. With the wide and rapid development of the social media platform, the rich social media data provide new opportunities for the prediction of movie box-office profits. Hence, it inspires us shed some of our obsession for causality in exchange for simple correlations. By letting us identify a suitable proxy for a phenomenon, correlations allow us to capture the present and predict the future. In this task, social media has the following advantages:

- Volumes of data about movies are available on social media. Movies are widely discussed on social media. According to our statistics, at least 10 million user posts talk about movies per week in the Sina micro blog. Therefore, sufficient data is available for the analysis.
- Data on movie box-office profits are easy to obtain. Income from the first week of a movie and its gross income can be acquired from the MTime in China and the Internet Movie Database (IMDB) in the US.

- Social media content and movie box-office profits have a clear logical correlation [6]. The user who posts a tweet to express his/her purchase intention for a specific movie indicates his/her interest in the movie and his/her likelihood to watch the movie. The first week pre-release data have the strongest correlation with the gross income than those in any other pre-release time periods [5]. After the movie's release, user posts, especially those with positive or negative sentiments, become a kind of electronic Word of Mouth. It can influence other potential customers [7] and further affect the gross income of the movie.

Using large-scale social media content, our approach seeks to predict movie box-office profits by mining correlation factors from unstructured texts. Most previous studies predicted the movie gross income based on structured IMDB data analysis of specific characteristics [1], e.g., the number of one-week-old theaters, the rating from the Motion Picture Association of America, director, main actors, movie's genre, budget, and so on, but with somewhat limited success. Nevertheless, recent work [8] has shown the power of social media in predicting financial market phenomenon such as stock price movement, product sales, and financial risk.

Rest of the paper is organized as follows: in section 2 describe the basic concepts of social media, in section 3 we describe areas where prediction with social media might be applied, provide a brief summary of the various features used in the prediction models are presented in section 4, section 5 provides list of the major metrics about social media used in prediction, in section 6 Previous research done in the area of predicting box office is presented briefly, the problem of predicting movie profitability or collection is given in section 7 and finally we conclude our paper with future research directions.

2. TECHNICAL BACKGROUND

In this section, we will introduce basic concepts about social media, and discuss their important characteristics.

2.1 Social network

A social network is a social structure comprising of persons or organizations, which usually are represented as nodes, together with social relations, which correspond to the links among nodes. The social relation could be both explicit, such as kinship and classmates, and implicit, for example friendship and common interest.

2.2 Social media

Social media comprise platforms to create and exchange user-generated content. Sometimes social media are called consumer-generated media (CGM). Social media are different from traditional media, such as newspaper, books, and television, in that almost anyone can publish and three access information inexpensively using social media. In contrast, traditional media (which is also referred as old media or legacy media) requires significant resources to publish contents. But social media and traditional media are not absolutely distinct. For example, major news channels have official accounts on Twitter and Facebook. There are many forms of social media that include blogs, social networking sites, virtual social worlds, collaborative projects, content communities and virtual game worlds. Some forms of social media lack a social network. Thus in blogspot.com, which is a famous blog platform, there is no social links among bloggers.

2.3 Social networking service

Social networking service is a set of online sites and applications, which at least consist of three parts: users, social links, and interactive communications. In fact, SNS is a subset of social media, which include the social network. On SNS, communication is interactive. For instance, for pure blogs, a non-SNS social media such as blogspot.com, the users' major motivations could be recording one's daily life, providing commentary and opinions, expressing feeling and emotion, demonstrating ideas via text, and keeping community. The first four motivations are all information sharing.

2.4 Why we need to predict automatically

Even though currently most predictions using social media can be done better by human agents, specifically experts, there are still good reasons for us to try to predict automatically.

3. PREDICTION SUBJECTS

In this section, we describe areas where prediction with social media may be made. Generally, a subject, that could be well predictable with social media, must meet the following requirements.

Firstly, the prediction subject must be human related event. On social media, users publish their opinions and beliefs. Prediction methods analyze, extract and integrate the information, and then according to the influence of persons to the predicted subject, make the prediction. But if the subject is non-human-related event, such as eclipse, even though there may be tons of users discuss this topic on social media; the users' thoughts have nothing to do with the development of that event. Consequently, the data on social media could not be used to predict natural events whose development is independent of human actions.

Secondly, if masses of people are involved, the distribution of composition of involved persons on social media should be the same as or similar to that in real world [9]. Because not everyone in real world will use social media, the users on social media could be treated as samples of the involved masses in most cases. But the sampling process is uncontrollable, which may lead to samples with built-in bias. Even though we cannot exclude biased samples completely, we should make sure the proportion of the biased samples is in the acceptable and reasonable range.

Lastly, the involved events should be easy to be talked in public. Otherwise, the contents on social media would be biased [9]. For example, three is social consensus that giving appropriate tips is good and excessively low tipping is impolite and unacceptable. Under such social pressure, almost nobody is willing to admit that he/she paid tips that were too low. The anonymous mode could be used in getting an answer to this issue but such an anonymous mode will have no information about relevant social network structures.

3.1 Marketing

There is some evidence that there is strong correlation between spikes in sale rank and the number of related blog posts. But at the same time, based on blog mentions, predicting whether tomorrow's sales rank for a particular item will be higher or lower than today's sales rank appears to be hard. There are two possible reasons for these seemingly contradictory conclusions. On one hand, there may be a delay between the increase of blog mentions and the increase of sale. On the other hand, the number of blog mentions may predict the change of sale. But the change of sale about one product does not necessarily change the sale rank of other products.

3.2 Movie box-office

To predict movie box-office with social media is one of the most studied area. In addition to the traditional prediction factors, such as MPAA rating and number of screens [1], social media contents could also be effective to predict box-office [6]. There are many reasons that predicting movie box-office a good subject for research.

3.3 Information dissemination

Information dissemination means how contents spread on the Internet. In other words, it refers to how users pay attention to different information. Since contents attract the users in an asymmetric way, the attention of most users is concentrated on a few contents [10]. Successful prediction could enhance the user experience by providing them with the most attractive information. Information dissemination could be researched on micro level and macro level.

3.4 Elections

Election prediction uses the survey of public opinion on political party or politician from a particular sample to predict the election result. Traditionally, the election polls could be done via telephone surveys. But thousands of calls easily lead to cost as high as tens of thousands of dollars. As a newly emerging method, web survey with social media provides an opportunity to do that with low cost.

3.5 Macroeconomic

The macroeconomic includes the regional, national or global economies. Some researchers are trying to use social media to deal with its trends, such as economic indices and stock markets. Generally the social media could not be used to accurately determine these trends alone, but could assist the researchers to capture or predict trends.

4. PREDICTIVE FEATURES

Accurate predictions depend on the predictive characteristics of the variables used in the model. While most prediction models proposed so far use the popularity at early moments as the only predictive variable there have been several attempts to include other features in a prediction model. We provide a brief summary of the various features used in the prediction models and report their predictive collection.

4.1 Characteristics of content creators:

The online media ecosystem is populated by content creators (independent producers, professional bloggers, mainstream mass media, or news agencies) with different but relatively stable – and maybe predictable – audience. Including the identity of the content creator in a prediction model is exploited by Bandari et al. who notice that the publisher of a news article is one of the strongest predictor of the number of tweets that a news article will generate [18].

4.2 Textual features:
Certain words or key phrases that probably refer to hot or controversial topics often produce a significant amount of attention. There have been two efforts to include textual features in a prediction model. Tsagkias et al. extract the top-100 most discriminative terms from various news sources and observe that these terms have a strong collection in predicting which articles will be highly commented [17]. Similarly, Marujo et al. show that popular key-phrases have a strong predictive power in predicting the number of views for news articles [19].

4.3 Content category:
Designing specialized prediction models depending on the category of the content showed little benefit in predicting the popularity of videos [20] and new articles [18]. The only notable exceptions have been signaled for YouTube Music videos [20] and news articles related to Technology section [18]. The low predictive collection of using this information in a prediction model can be explained by the overlapping scope of categories, with content often belonging to multiple categories at once [18].

4.4 Named entity identification:
Popular entities in the real world (people, locations, or organizations) can often be a catalyst of user attention in the online sphere. Tsagkias et al. observe a strong influence in including popular entities from Netherlands in a prediction model designed to spot news articles that will receive a high number of comments [17].

4.5 Sentiment analysis:
The specific emotion triggered by web content is highly correlated with its online popularity but extracting the correct sentiment and learning how to use this information for popularity prediction is a difficult task. The subjectivity of the language has shown little predictive power in predicting the volume of tweets for online news stories [18]. However, it has been observed that articles that are written in a more positive or negative voice, associated with strong emotions (e.g., admiration or anger), are good indicators of how viral articles will become. In addition, Oghina et al. observe that subjective terms from the discussions about movies on Twitter can successfully be used as a predictive variable in predicting movie ratings on IMDb [15].

4.6 Social media signals:
As we saw in Section 7.2, social media conveys valuable information about web content popularity. Castillo et al. show that the attention that news articles generate across social networks (number of Facebook shares, number of tweets and retweets, and the language of the Twitter messages) is effective in predicting the popularity of articles on a news site. Oghina et al. successfully use information from Twitter and YouTube to predict movie ratings on IMDb. Another example of the predictive power of social media has been reported by Roy et al. who show that the popularity of a topic on Twitter provides a good indication that a YouTube video will experience a sudden burst in popularity [22].

4.7 Social sharing viewing behavior:
Yahoo! Zync is an application that allows users to share and jointly manipulate video content in real time. Shamma et al. study how users' actions during a sharing session can be used to predict the popularity of YouTube videos and observe that these interactions are strong indicators of videos with a high number of views [23].

4.8 Real-world features:
Content published in online media is strongly related to real-world events but transferring information from the physical to the online world is very challenging. An attempt to employ real-world information in the predictions process has been done by Tsagkias et al. who show that there is an insignificant benefit in using the weather conditions (average temperature in Netherlands) to predict the number of comments for news articles [17].

5. THE PREDICTORS
In this section, we will list the major metrics about social media used in prediction. Mostly these metrics alone do not have sufficient prediction power but their combinations work better. These predictors may be divided into two categories: message characteristics and social network characteristics.

5.1 Message characteristics
Message characteristics focus on the messages themselves, such as the sentiment and time series metrics. If the research focus on general objects, all the available posts are fetched with timestamps. Otherwise, the search result with man-crafted keywords is preferred.

(a) Sentiment metrics

The sentiment metrics are the static features of posts. In addition to the general sentiments discussed in the following, there are some specific sentiment categories, such as happiness and anxiety, on a case-by-case basis. Because they lack generality, we do not investigate them in detail. But the concept, extraction and usage of them are same as these of general ones.

(b) Time series metrics

Time series metrics try to investigate the posts dynamically, including the speed and process of the message generation. The posts generating rate means how quickly the messages are produced.

5.2 Social network characteristics

Social network characteristics measure structure features. We also call these characteristics as metrics/measures in social network analysis. Being long studied, these are so many characteristics that it's impossible to list and investigate all of them here. So we just enumerate and briefly discuss the most used ones in predictions.

6. LITERATURE SURVEY

Previous research done in the area of predicting box office success has applied different techniques such as neural networks [2], statistical Baysian [24] and linear regression modeling techniques [25]. The most recent work in the literature in box office prediction using neural networks is by [6]. They use a multilayer perceptron with a Levenberg-Marquant learning algorithm to classify Bollywood films into 3 categories based on the box office collection using 111 instances of observations and 8 attributes. It resulted in a classification accuracy of 93.1%. A multilayer perception is used in the work done by Li, Jianhua, and Suying, [21]. They use the data of 241 Chinese/American films and attempt to predict each film into 6 classes based on box office success ranging from a box office blob to bomb or mega success. based on 11 weighted continuous variables. The model used a multilayer back propagation network consisting of 30 nodes in the first hidden layer and 10 nodes in the second hidden layer, which was then standardized and validated through a 6-fold cross validation to measure collection. The results were not very accurate at classifying movies in their actual class but had high relative accuracy (97.1%) indicating that the predicted class was one class away from the actual class.

A similar study is done by Delen and Shard [1]. They use 834 movies from 1998-2002 and run it through a multilayer perceptron network. They try to classify each movie into one of 9 classes based on its box office profit using 7 continuous variables and a 10-fold cross validation. Accuracy is measured by the percentage correct classification rate and the 1-away classification rate which resulted in a 36.9% accuracy rate and a 75.1% accuracy rate respectively. The work compares the neural network approach to other statistical analysis strategies such as the discriminant analysis, multiple logistic regression and decision trees. However, the neural network approach proved to be the most accurate.

Chang and Lee use a Bayesian Belief Network to determine the causal relationships between 18 variables in predicting box office success for Korean movies [24]. Sensitivity analysis is used to determine the important attributes. They use the number of movie goers as their metric for success and divide them into two groups. The first group is further split into inferior and superior categories based on the median and the second group is split into bad, standard and excellent categories. Movies in the superior and excellent category are the most successful in the two groups. When compared to artificial neural networks and decision tree approaches, the Bayesian Belief Network proves to be the most accurate.

None of the above papers account for the social media data which has become an important indicator and influencing factor of a movie being successful. Skiena and Zhang use ratings from the IMDB, a popular database of information about movies, and sentiment analysis on news data gathered by "Lydia" (www.textmap.com) to determine the gross profit of a movie.

They use linear regression and K nearest neighbor models and demonstrate that the linear regression model is better at predicting low gross turnover movies while the K-nearest network is good in predicting high gross turnover movies. Accuracy of the approach increases when the ratings from IMDB are combined with the news analysis [25].

Breuss, de Rijke, Oghina, and Tsagkias [15] explore the idea that user activity on social media may have an influence on box office collection of a movie. They look for a correlation between the movie information extracted from the IMDB and people's remarks regarding the extracted movies on social media websites such as the Twitter and Youtube. More specifically the authors look into the qualitative data (textual) which represent the comments of people about the quality of a movie, and the quantitative (surface) data which represent the number of people commenting about a movie. These social media data is used in a linear regression model and the results indicate that the Youtube like to dislike ratio and the textual analysis of Twitter data are the best in predicating the IMDB rating of a movie [2].

Asur and Huberman research the effect of Twitter on box office profits [6]. Extracting data from the Twitter Search API, they identify the week before the movie is released as the critical period. They measure the rate of tweets and compare it to box office profit using linear regression. These two variables show a strong positive correlation. When compared to the "Hollywood Stock Exchange" which uses prices for "movie stocks" to predict the profit, the Twitter model surpasses it in accuracy. It also shows better accuracy than the news/IMDB model of Skiena and Zhang [25]. Sentiment analysis further improves the model by a significant margin.

This paper proposes a decision support system to aid movie investment decisions at the early stage of movie productions. The system predicts the success of a movie based on its profitability by leveraging historical data from various sources. Using social network analysis and text mining techniques, the system automatically extracts several groups of features, including "who" are on the cast, "what" a movie is about, "when" a movie will be released, as well as "hybrid" features that match "who" with "what", and "when" with "what". Experiment results with movies during an 11-year period showed that the system outperforms benchmark methods by a large margin in predicting movie profitability. Novel features we proposed also made great contributions to the prediction. In addition to designing a decision support system with practical utilities, our analysis of key factors for movie profitability may also have implications for theoretical research on team collection and the success of creative work [11].

In the past few years, social media have seen its presence and influence all over the globe. It has become an important tool for socially networking and sharing of content. And still, the data spread across these different WebPages remains hugely unexploited. In this paper, we exhibit how social media content may be used to anticipate motion picture outcomes. In general, we used the traditional data from social media platforms such as Face book and Twitter to forecast box-office profits and its popularity for movies. The paper gives an insight on procuring those comments or posts from social media and performs a semantic analysis to see if a particular movie is getting the exposure or not. We show that extracting the features and contents of social media gives us a hope to discover social structure attributes, study action patterns qualitatively and quantitatively. We also show how sentiments mined from these platforms may be used to improve the anticipation power of our social media [12].

Forecasting the box office success of upcoming movies is an important task for the entertainment industry, and is inherently complex due to its extremely unpredictable nature. Prior work has used Twitter data analysis to predict the success. However, the noisy nature of Twitter texts rendered them unreliable. In this work, we proposed a transfer learning approach to overcome the issues related to tweets. To accomplish this, we made use of the much cleaner and more descriptive online reviews of already released movies. In the proposed approach, we performed knowledge transfer by learning a common feature representation that helped to reduce the divergence between the reviews and the Twitter data. This enabled us to train a more accurate classifier on the transformed reviews data and employ it for predicting on the transformed tweets [13].

Predicting a movie's opening success is a difficult problem, since it does not always depend on its quality only. External factors such as competing movies, time of the year and even weather influence the success as these factors influence the Box Office sales for the moving opening. Nevertheless, predicting a movie's opening success in terms of Box Office ticket sales is essential for a movie studio, in order to plan its cost and make the work profitable. I introduce a simple solution for predicting movie success in terms of financial success and viewer reacceptance. As a result, this approach achieved decent estimations, allowing theatre planning to a certain extent, even for small studios [14]

We predict IMDB movie ratings and consider two sets of features: surface and textual features. For the latter, we assume that no social media signal is isolated and use data from multiple channels that are linked to a particular movie, such as tweets from Twitter and comments from YouTube. We extract textual features from each channel to use in our prediction model and we explore whether data from either of these channels can help to extract a better set of textual feature for prediction. Our best performing model is able to rate movies very close to the observed values. [15]

In recent years, social media has become ubiquitous and important for social networking and content sharing. And yet, the content that is generated from these websites remains largely untapped. In this paper, we demonstrate how social media content can be used to predict real-world outcomes. In particular, we use the chatter from Twitter.com to forecast box-office profits for movies. We show that a simple model built from the rate at which tweets are created about particular topics can outperform market-based predictors. We further demonstrate how sentiments extracted from Twitter can be utilized to improve the forecasting power of social media [6].

Social media has become pervasive and important for social networking and content sharing but then, the substance that is created from these sites remains to a great extent undiscovered. In this paper, we exhibit how online networking substance can be utilized to anticipate genuine results. System we propose will make use of Twitter Auth for streaming tweets. Tweets will be in unstructured format so our system will bring unstructured data into structured format for sentiment analysis. Once data brought into structured format, system will apply weights to tweets depending upon various criteria such as followers, following of actor, actress, director, producer and also rate of tweets on film hash tag. We further analyze sentiments extracted from Twitter which can be further utilized for forecasting movie box office collection of first week and overall collection [16]

7. PROBLEM STATEMENT

Previous research has tried to solve this problem of predicting movie profitability or collection; however it does sub-optimally by choosing traditional factors such as movie script [1], advertisement budget, number of opening theaters, production house, MPAA Rating [2], star cast [3], etc. A number of interesting lines of recent work have pursued the pre-release financial movie success problem using a variety of social media data such as Google and YouTube trailer search volume [5], critic reviews [6], tweets [8], Wikipedia activity level [4]. However, the collection of the models in [4] significantly degraded for movies, which had medium and low popularity. We attribute such degradation in the collection of the model to the choice of social media signal made in the above-mentioned works.

Thus, identification of signals, which could be equally relevant to all profitability classes of movies in order to derive classification or predictive models, remains an open problem [4].

Upon investigation, we observe that current research suffers from the following limitations:

- Social media signals such as tweet, blogs and reviews lack worldwide applicability due to language barrier.
- Publicly available meta-data on social multimedia sites like YouTube still remains unexplored in addressing the problem.
- Star popularity in influencing the collection of movies still stands in debate.
- A comparative model to compare the relevance of socially generated meta-data from media and multimedia sites in solving the problem remains absent.

We formulate two concrete tasks that help us understand the problem scenario in a better way.

- Determine whether the social media signals we use improve classification of movies by profitability.
- Do a comparative analysis to test the amount of relevance that different social media signals have on the collection of movies

CONCLUSIONS

A movie success does not depend only on those features related to movies. The number of audience plays a vital role for a movie to become successful. Because the whole point is about viewers, the entire industry will make no sense if there is no audience to watch a movie. The number of tickets sold during a specific year can indicate the number of viewers of that year. And the role of movie audience depends on many situations like political conditions and economic stability of a country. As discussed, the easy availability of APIs provided by Twitter, Facebook and News services has led to an 'explosion' of data services and software tools for scraping and sentiment analysis, and social media analytics platforms. This paper is written for (social networking) researchers who seeking to analyze prediction of movies via of social media. It presents a comprehensive review of social media analytics for social networking, wikis, really simple syndication feeds, blogs, newsgroups, chat and news feeds etc.

REFERENCES

[1.]Sharda, Ramesh, and Dursun Delen. "Predicting box-office success of motion pictures with neural networks." Expert Systems with Applications 30, no. 2 (2006): 243-254.
[2.]Ghiassi, M., David Lio, and Brian Moon. "Pre-production forecasting of movie profits with a dynamic artificial neural network." Expert Systems with Applications 42, no. 6 (2015): 3176-3193.
[3.]Zhang, Li, Jianhua Luo, and Suying Yang. "Forecasting box office profit of movies with BP neural network." Expert Systems with Applications 36, no. 3 (2009): 6580-6587.
[4.]Ivasic-Kos, Marina, Miran Pobar, and Luka Mikec. "Movie posters classification into genres based on low-level features." In Information and Communication Technology, Electronics and Microelectronics (MIPRO), 2014 37th International Convention on, pp. 1198-1203. IEEE, 2014.
[5.]Zhang, Wenbin, and Steven Skiena. "Improving movie gross prediction through news analysis." In Web Intelligence and Intelligent Agent Technologies, 2009. WI-IAT'09. IEEE/WIC/ACM International Joint Conferences on, vol. 1, pp. 301-304. IEEE, 2009.
[6.]Asur, Sitaram, and Bernardo A. Huberman. "Predicting the future with social media." In Proceedings of the 2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology-Volume 01, pp. 492-499. IEEE Computer Society, 2010.
[7.]Lica, Liviu, and Mihaela Tuta. "Predicting product collection with social media." Informatica Economica 15, no. 2 (2011): 46.
[8.]Bollen, Johan, Huina Mao, and Xiaojun Zeng. "Twitter mood predicts the stock market." Journal of computational science2, no. 1 (2011): 1-8.
[9.]D. Huff. How to Lie with Statistics. WW Norton & Company, 1993, pp. 13-28.
[10.]F. Wu and B. A. Huberman, "Novelty and collective attention.," Proceedings of the National Academy of Sciences of the United States of America, vol. 104, no. 45, pp. 17599-601, Nov. 2007.

[11.] Lash, Michael T., and Kang Zhao. "Early predictions of movie success: The who, what, and when of profitability." Journal of Management Information Systems 33, no. 3 (2016): 874-903.
[12.] Parth Bhavsar, Adheesh Kumar, Sumesh Kumar , Amit Gaur , S.G.Sheikh, “ ”, International Research Journal of Engineering and Technology (IRJET), Volume: 04 Issue: 05 | May -2017.
[13.] Catherine, Rose, and Sneha Chaudhari. "Predicting Movie Success from Tweets."
[14.] BACHELORARBEIT, “Predicting Movie Success”, to the Faculty of Informatic at the Vienna University of Technology.
[15.] Oghina, Andrei, Mathias Breuss, Manos Tsagkias, and Maarten de Rijke. "Predicting imdb movie ratings using social media." In European Conference on Information Retrieval, pp. 503-507. Springer, Berlin, Heidelberg, 2012.
[16.] Mulay, Snehal A., Shrijeet J. Joshi, Mohit R. Shaha, Hrishikesh V. Vibhute, and Mahesh P. Panaskar. "Sentiment Analysis and Opinion Mining With Social Networking for Predicting Box Office Collection of Movie." International Journal of Emerging Research in Management & Technology(2016): 74-79.
[17.] Bandari, Roja, Sitaram Asur, and Bernardo A. Huberman. "The pulse of news in social media: Forecasting popularity." ICWSM 12 (2012): 26-33.
[18.] Tsagkias, Manos, Wouter Weerkamp, and Maarten De Rijke. "Predicting the volume of comments on online news stories." In Proceedings of the 18th ACM conference on Information and knowledge management, pp. 1765-1768. ACM, 2009.
[19.] Marujo, Luís, Miguel Bugalho, João Paulo da Silva Neto, Anatole Gershman, and Jaime Carbonell. "Hourly traffic prediction of news stories." arXiv preprint arXiv:1306.4608(2013).
[20.] Pinto, Henrique, Jussara M. Almeida, and Marcos A. Gonçalves. "Using early view patterns to predict the popularity of youtube videos." In Proceedings of the sixth ACM international conference on Web search and data mining, pp. 365-374. ACM, 2013.
[21.] Zhang, Li, Jianhua Luo, and Suying Yang. "Forecasting box office profit of movies with BP neural network." Expert Systems with Applications 36, no. 3 (2009): 6580-6587.
[22.] Roy, Suman Deb, Tao Mei, Wenjun Zeng, and Shipeng Li. "Towards cross-domain learning for social video popularity prediction." IEEE Transactions on multimedia 15, no. 6 (2013): 1255-1267.
[23.] Shamma, David A., Jude Yew, Lyndon Kennedy, and Elizabeth F. Churchill. "Viral Actions: Predicting Video View Counts Using Synchronous Sharing Behaviors." In ICWSM. 2011.
[24.] Amato, Filippo, Alberto López, Eladia María Peña-Méndez, Petr Vaňhara, Aleš Hampl, and Josef Havel. "Artificial neural networks in medical diagnosis." (2013): 47-58.
[25.] Zhang, Wenbin, and Steven Skiena. "Improving movie gross prediction through news analysis." In Web Intelligence and Intelligent Agent Technologies, 2009. WI-IAT'09. IEEE/WIC/ACM International Joint Conferences on, vol. 1, pp. 301-304. IEEE, 2009.