RESEARCH ARTICLE

PREDICTION USING SOCIAL MEDIA

*Gargi Patil and Tarak Patel

School of Computer Science and Engineering, Vellore Institute of Technology University, Vellore-632 014, Tamil Nadu, India

---

ARTICLE INFO

ABSTRACT

Social media consists of lot of user generated data. It contains opportunity for creation, publication and transfer of ideas and thoughts which is considered as user generated data. The past few years have laid to large amount of growth in social media. Social networks are affecting the way how human interact with each other and share different perceptions. The data generated by user contains large amount of unprocessed information related to the perception and feelings of user towards a particular entity. In this paper, we discuss the importance of data generated by social media in order to predict about future events.

INTRODUCTION

Human beings are very much interested in knowing about their future, especially when it is related to their personal life or to their personal interests and business. Sometimes we relate human brain abilities with an ability to predict future affairs.In this paper we are providing a detailed explanation about an intelligent system which can help us to predict the future events on the basis of social network data. We are concentrating on a particular service and various companies providing the same service will be evaluated on the basis of data available on social network and the future status of company in market will we predicted.

Literature Survey

Schoen H *et al*. (2013) shown the issues faced when trying to make predictions using social media. Author here introduced a taxonomy of all models that is being used in the past time to predict future events with the help of social media data like survey , markets and statistical models. For every model discussed its advantages and problems faced during applying it had been discussed. Further author focused on four areas :influenzaincidence, product sales, stock market movement, and electoral results , of prediction using social media that had caught the eyes of of researchers. Finally they introduced accepted papers for this issue and discuss their high level of contributions. Javeed *et al*. (2014) tackled the problem faced during information summarization form big textual data, also corporating the sentiment towards the information.

---

*Corresponding author: **Gargi Patil,**
School of Computer Science and Engineering, Vellore Institute of Technology University, Vellore-632 014, Tamil Nadu, India

For this purpose they depicted new concept know as sentimental causal rules which is combination of sentiment analysis and causal rules. They also proposed a method that extracts sentimental causal rules from data sources of text kind. Also authors presented a case study of *Kurdish issue* in *Turkey* as the main topic and Twitter as the main source of data to extract sentimental causal rules. The methodology proposed here is automatic processing rather than manual, of tweets which is hard task due to syntax mistake, junk words. Romero *et al*. (2015) reviewed 15 web services functionality related to sentimental Analysis. Some of the services were of private companies but allowed free access to functionalities they provide. Using these services like Alchemy and Semantria the results obtained could be used for any kind of text sentiment analysis. It is helpful for users with quite large texts and want to classify them as negative or positive. Tools like Musicmetric and Uclassify could be considered.

If the text has ironic sentences then these tools can be good options. On the other side, authors also discard some tools like Wingify or Viralheat because of results obtained. As a final statement authors concluded that these tools suffers problems of excessive simplicity during classification or lack of capabilites to combine rating froms different types of sentences or paragraphs to get a generalized rating about a opinion. Tripath *et al*. (2015) attempted to make classify sentiment analysis for movie review using techniques of machine learning. Two of the most know approach: Naive Bayes(NB) and Support Vector Machine (SVM) were discussed. The results of this approaches were compared with other approaches discuss in other papers too and it was found that SVM classifier beats every other classifier in terms of

prediction of sentiment of a review. Thelwall *et al*. (2009) concluded that the use of multiple classifiers in a hybrid manner produces a better and effective result in terms of micro and macro-averaged F1 than any other individual classifier. With the help of sentiment Analysis Tool (SAT), semi-automatic , complementary approach can be apply that is each individual classifier contributes to other individual classifiers to obtain a better result and high level of effectiveness. Also author claimed that induction algorithm can be used to generate a collection of induced antecedents which are too sparse for deeper level of analysis therefore for real world scenario, it is better to have two sets of rule, the original one and the induced rule set.

In (Shahana, 2015) shahana PH *et al*. discusses the problem faced in sentimental classification. They found that the best method to extract sentiment from review is unigram. It was found that unigram with stemming with stop word gives accuracy of 82.9% as compared to unigram with stemming without stop word with 83% in positive class. For negative class unigram for both method gives 83.1% accuracy. Also author concluded that with the help of information gain, both classes give better result. Barakos, (?) presents an overview of Social Media based forecasting , hidden potential of innovative development product and methods for data processing purpose as well as criticism , limitation were also outlined.

With this a new frameworks was developed customized for the predicting future trends of market and needs of customers. Various tools were outlined which can be used for prediction purpose using social media data. Wang *et al*. (2016) started by taking efficient market hypothesis in which they comments on stock using social media like twitter and then preprocessing the obtained data to emotion vectors. After the after calculation of segment value of each and every stock's emotion vector, they found that segment value obtained is very sensitive to movement of stock price. In addition to this, they found that SVM model with segment index has higher accuracy than compared to SVM model without segment index.

METHODS

A basic question which must be addressed so as to deal with the problem statement is: What factor makesthe prediction on the basis of social media data possible? From this we can easily say that, the result i.e. prediction itself has to be present within the data. Another thing is, the collection of data should maintain the encoding of the result.

Third thing is, the evaluationcarried out on the gathered data should be able to disclose the prediction. These three things are very important in order to get the proper prediction. In the method which we have used, we are mainly concerned about the message characteristics. After fetching the data from social network the sentimental metrics will be calculated for further processing the data.

This method includes various steps like

- Creating the authorized access to the social network in order to fetch the data.

- Fetching the particular amount of data on the basis of the requirement. (Test sample)
- Processing the row data in order to find sentimental metrics.
- It involves calculation of positive and negative sentiments.
- It involves the procedure to check for neutral sentiments.
- On the basis of sentimental metrics judging the individual elements (Here in this case the company providing the same service).
- On the basis of individual results doing comparative analysis.
- Based on the comparative analysis predicting the future market position of company.
- Presenting the data in processed manner. (Graph, Result variables, etc.)

Here the statistical analysis includes the statistical representation of processed data and inferring results from it. The method is explained in Figure 1. The data processing involves fetching the data from the social media. The extracted data is then cleaned by removing unnecessary and unimportant parts of data like stop-words, punctuation marks and blank spaces. The cleaned data is then processed in order to separate the words in each comment. The comment is then processed by comparing each word in comment to the standard set of positive and negative words. The number of matches are recorded and the score is calculated. The general formula for calculating the score is,

Score= Positive score- Negative score

Here, positive score refers to the number of positive matches and the negative score refers to the number of negative matches. The train set we are using is the set of standard positive and negative words. We import this data in the system in order to train the system. Then it is applied on the test set which is in this case; the data from social media.

- If the Score>0 then the comment is considered as positive.
- If the Score<0 then the comment is considered as negative.
- If the Score=0 then the comment is considered as Neutral.

All the comments are processed by the given method and the number of positive, negative and neutral comments is calculated. The entity having more positive comments is considered to have better performance in future.

The entity having more negative comments is considered to have bad performance in future. The entity having more neutral comments is considered to have static performance in future.

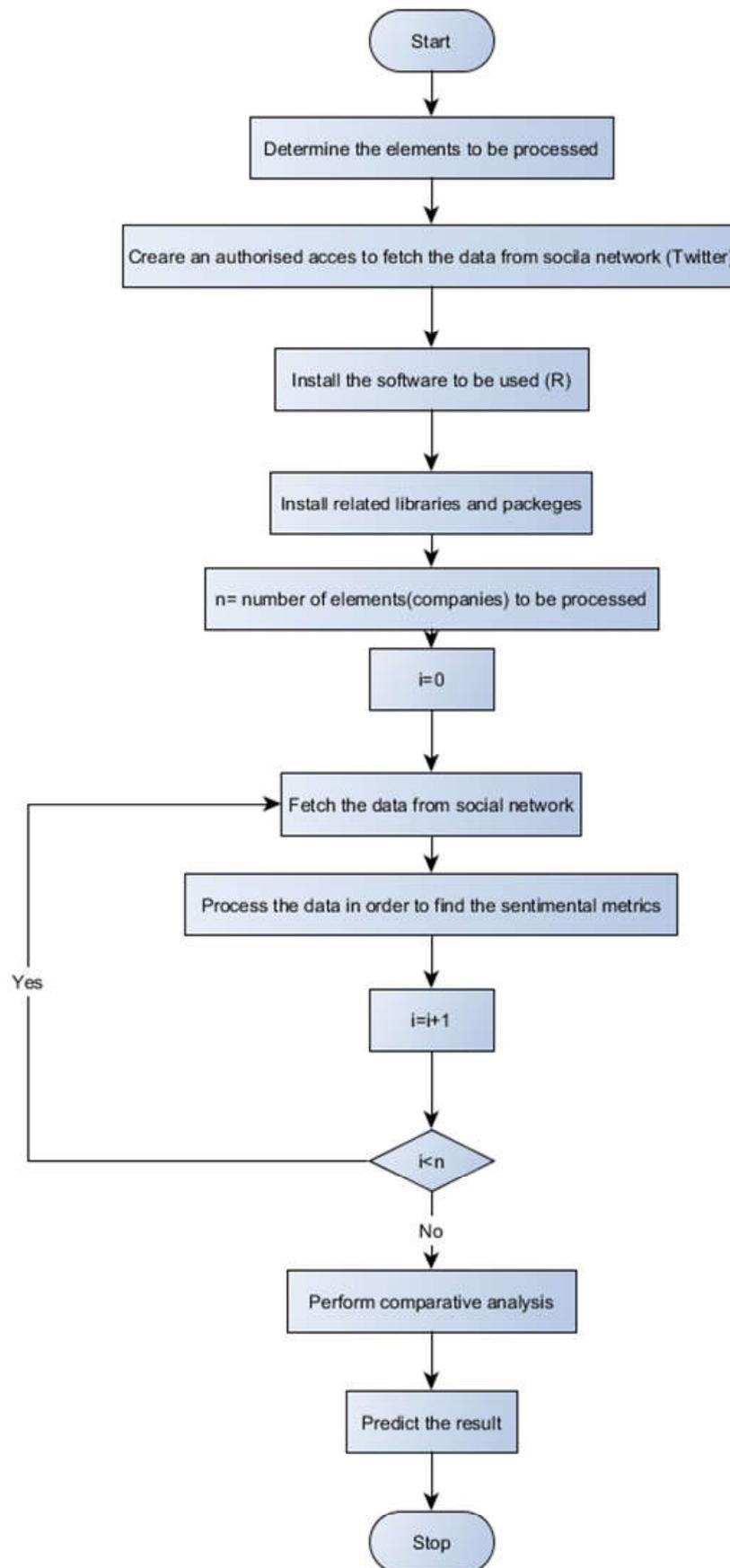The comparative analysis is done between the entities lead to the prediction.

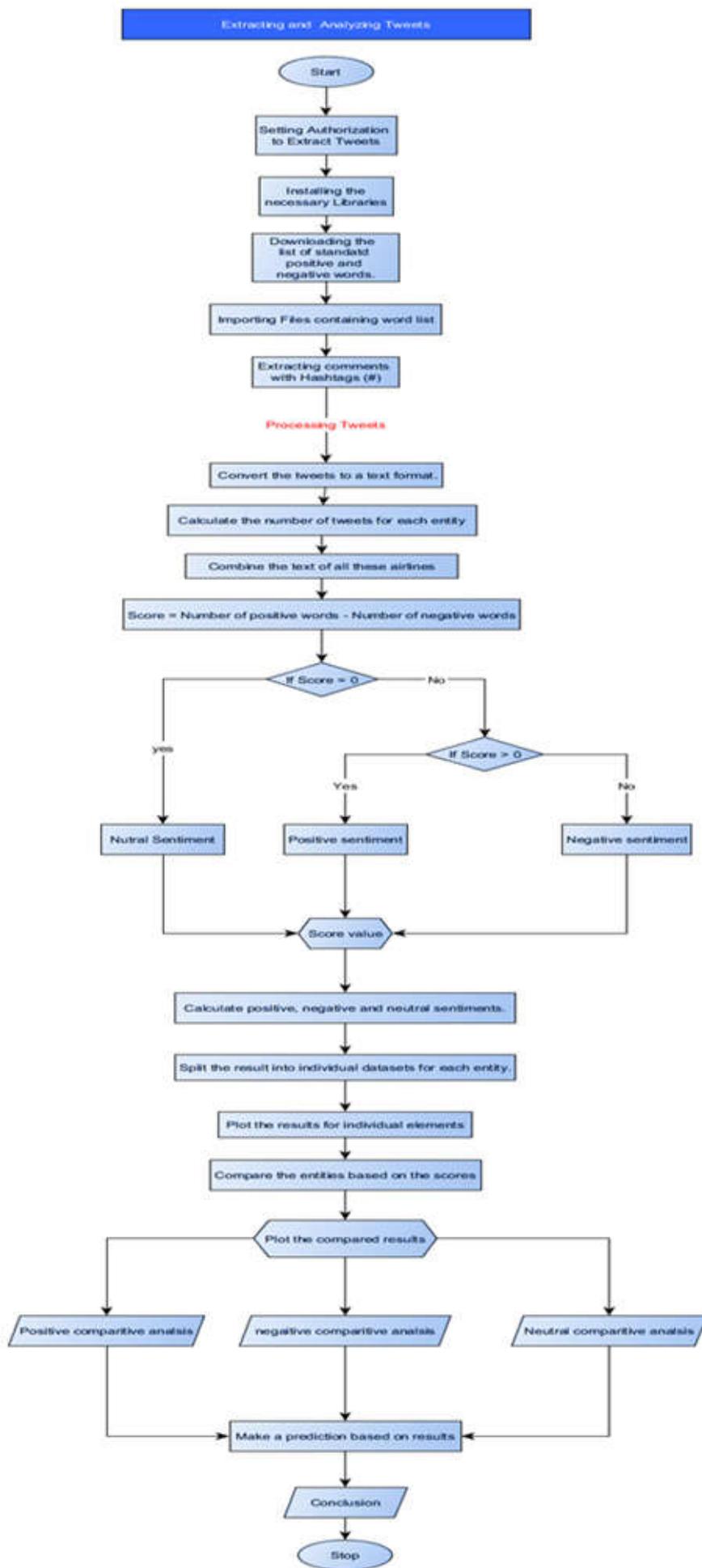**Figure 1. Flowchart- Implementation of the project**

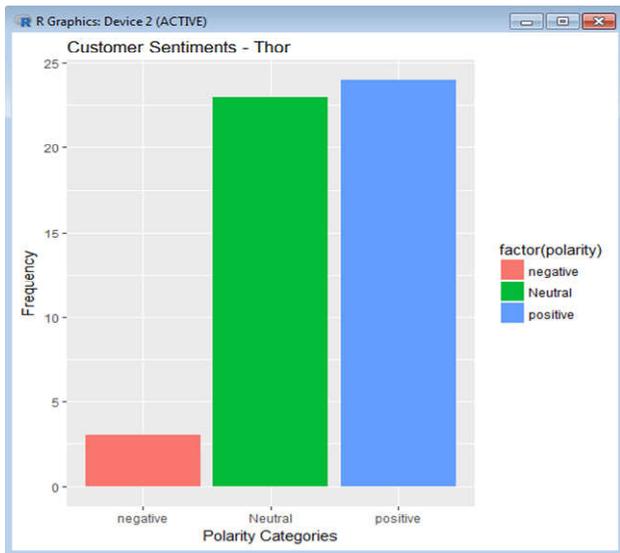**Figure 2. Flowchart- Extraction and analysing the data**

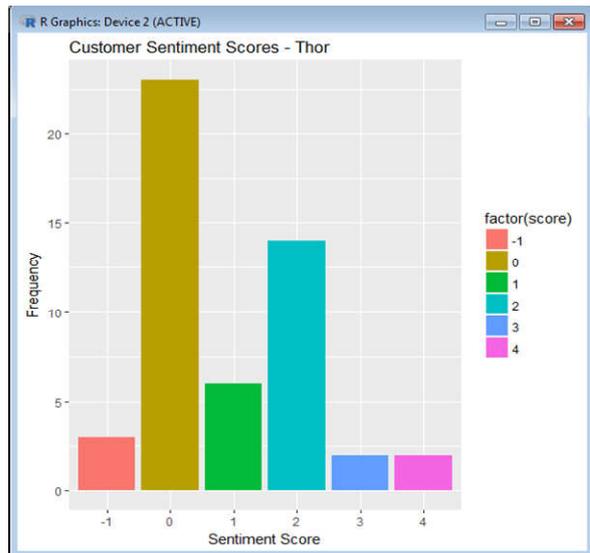**Figure 3.1. Thor movie related polarity analysis**

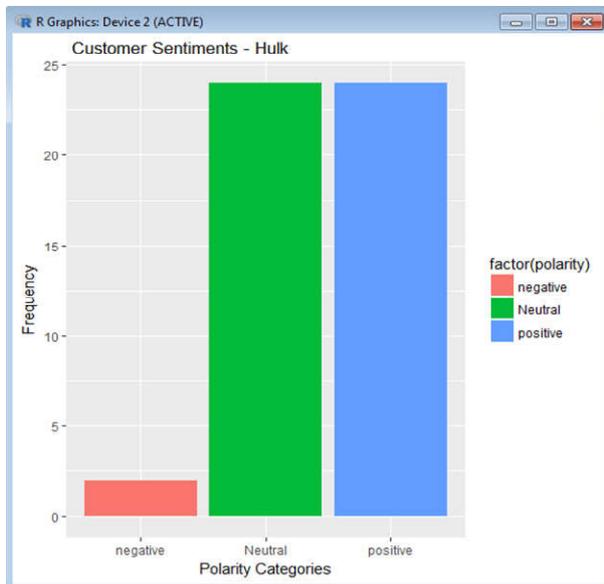**Figure 3.2. Thor movie related score analysis**

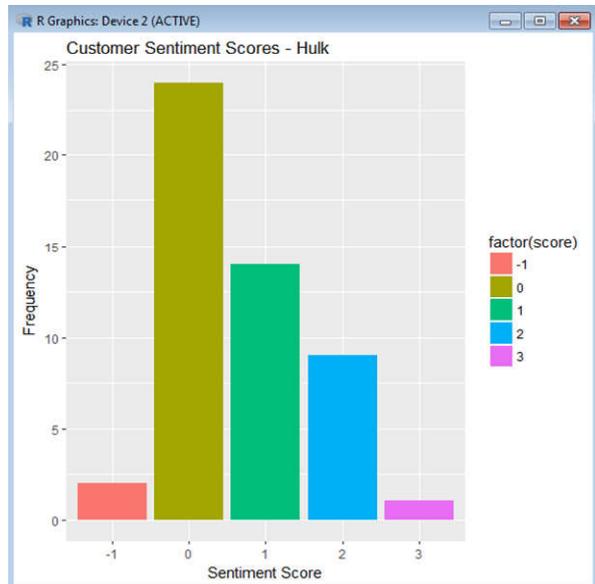**Figure 4.1. 'Hulk' movie related polarity analysis**

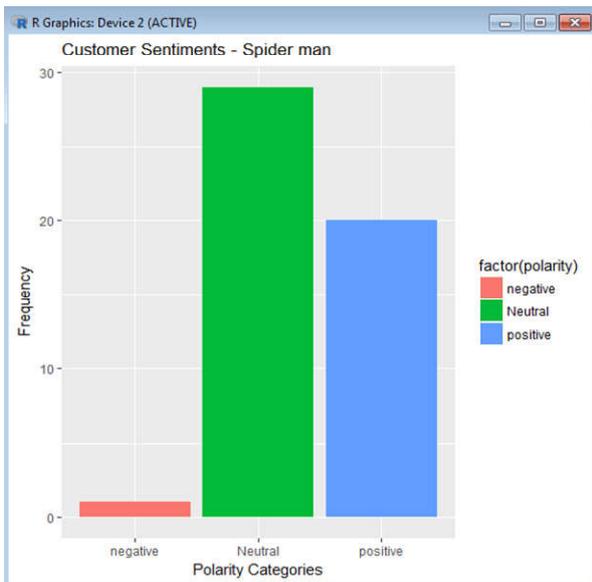**Figure 4.2. 'Hulk' movie related score analysis**

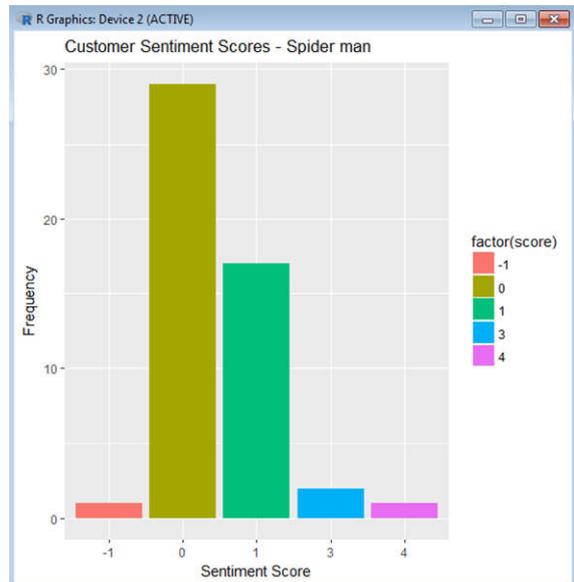**Figure 5.1: 'Spider Man' Movie related polarity analysis**

Figure 5.2: 'Spider Man' Movie related score analysis

According to the analysis carried on the data related to the 'Thor' Movie the positive score is greater. Negative score is very less. Neutral is also greater. Hence the total score can be considered as positive. According to the analysis carried on the data related to the 'Hulk' Movie the positive score is greater. Negative score is very less. Neutral is also greater. Hence the total score can be considered as positive.
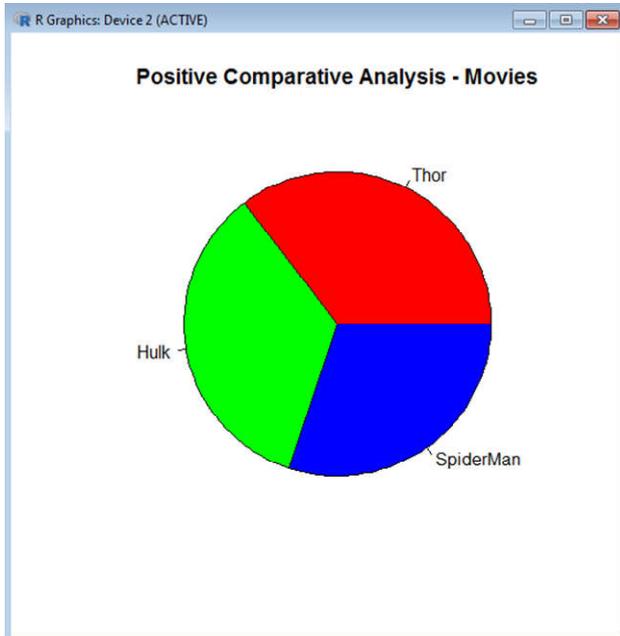

Figure 6.1. Positive comparative analysis

According to the analysis carried on the data related to the 'Spider Man' Movie the positive score is greater. Negative score is very less. Neutral is also greater. Hence the total score can be considered as positive. While doing the comparative analysis the results can be visualised using pi charts. For better understanding of the results the pi charts are divided into three types: first is the pi charts representing the positive count for each entity, second is the pi charts representing the negative count for each entity, third is the pi charts representing the neutral count for each entity.
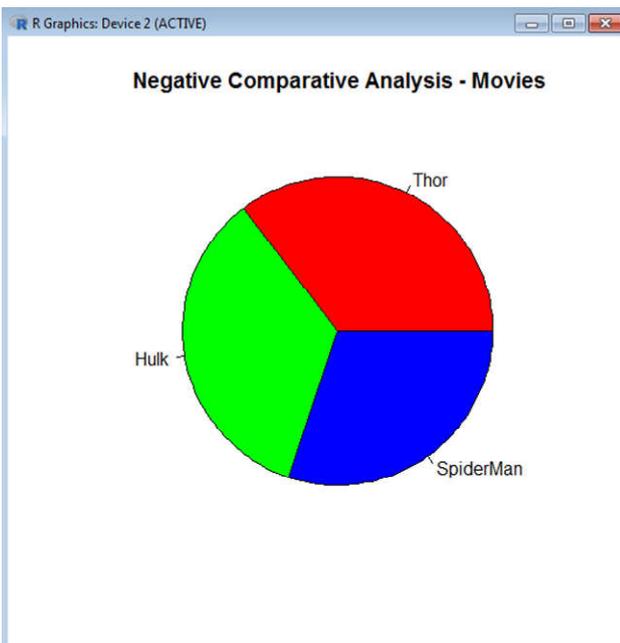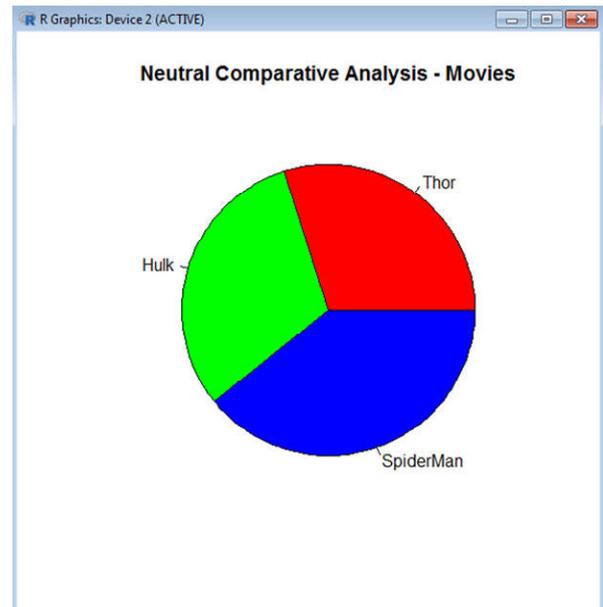

Figure 6.2. Negative comparative analysis

Figure 6.3. Neutral comparative analysis

All of the tree positive, negative and neutral comparative analysis shows that the movie series popularities are nearly same for all the three movies. Hence if all the movies are released simultaneously then it may lead to customer unsatisfaction as the users are interested in all the movies. This may lead to loss for the Marvel productions.

RESULTS

Presently Social media makesa great amount of data available, which can be used for research purpose by data scientists. The data is usually about users and their social interactions; henceit offerseconomists,computer scientists and social scientists, and statisticians lot of opportunities for research. The implemented system can be effectively used to guess the future marker risks by the companies. The system can be used by consumers in order to better analyse where to spend their money in order to bet better outcomes. The companies can guess about the competitions in market and take action in order to keep the growth rate constant. This predictive analysis enables companies to understand consumer sentiments in relation to specific service. These results can also be used to improve the products and services. In the long run, this predictive analysis can help business enterprises in improving the overall market position and propel business growth.

Applications: This system with few changes can also be used for:

Product Sales

It is another area with substantial research is that of predicting product sales, such as books, video games, and movie tickets.

Stock Market Movement

The stock market is, in some sense, one of the more promising areas to apply statistical modelling and, indeed, there is a rather large body of work regarding its predictability using social media data.

Electoral Results

Predicting election outcomes from social media, in particular using Social network data

REFERENCES

Barakos, M. *Social Media and Forecasting: What is the potential of Social Media as a forecasting tool?* (Bachelor's thesis, University of Twente).

Dehkharghani R, Mercan H, Javeed A, Saygin Y. 2014. Sentimental causal rule discovery from Twitter. Expert Systems with Applications. Aug 31;41(10):4950-8.

Goel S, Goldstein DG. 2013. Predicting individual behavior with social networks. Marketing Science. Oct 24;33(1):82-93.

Pang B, Lee L. 2008. Opinion mining and sentiment analysis. Foundations and Trends® in Information Retrieval. Jul 7;2(1–2):1-35.

Prabowo R, Thelwall M. 2009. Sentiment analysis: A combined approach. *Journal of Informetrics.* Apr 30;3(2):143-57.

Schoen H, Gayo-Avello D, Takis Metaxas P, Mustafaraj E, Strohmaier M, Gloor P. 2013. The power of prediction with social media. *Internet Research.* Oct 14;23(5):528-43.

Schoen H, Gayo-Avello D, Takis Metaxas P, Mustafaraj E, Strohmaier M, Gloor P. 2013. The power of prediction with social media. *Internet Research*. Oct 14;23(5):528-43.

Serrano-Guerrero J, Olivas JA, Romero FP, 2015. Herrera-Viedma E. Sentiment analysis: A review and comparative analysis of web services. *Information Sciences*. Aug 1;311:18-38.

Shahana PH, Omman B. 2015. Evaluation of Features on Sentimental Analysis. Procedia Computer Science. Jan 1;46:1585-92.

Tripathy A, Agrawal A, Rath SK. 2015. Classi詮갱ation of Sentimental Reviews Using Machine Learning Techniques. Procedia Computer Science. Jan 1;57:821-9.

Wang Y, Wang Y. 2016. Using social media mining technology to assist in price prediction of stock market. InBig Data Analysis (ICBDA), 2016 IEEE International Conference on Mar 12 (pp. 1-4). IEEE.

Yu S, Kak S. 2012. A survey of prediction using social media. arXiv preprint arXiv:1203.1647. Mar 7.

****