

# Outcome and Prediction of Popularity of Motion Picture Using Social Media

Parth Bhavsar<sup>1</sup>, Adheesh Kumar<sup>2</sup>, Sumesh Kumar<sup>3</sup>, Amit Gaur<sup>4</sup>, S.G.Sheikh<sup>5</sup>

<sup>1,2,3,4</sup> Student, Department of Computer Engineering, Sinhgad Institute of Technology, Lonavala, Maharashtra, India

<sup>5</sup> Ass. Professor, Department of Computer Engineering, Sinhgad Institute of Technology, Lonavala, Maharashtra, India

\*\*\*

**Abstract** - 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 Facebook and Twitter to forecast box-office revenues 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.

**Key Words:** Sign Language, Simple Sentence, Natural Language, Image Processing, Speech Recognition.

## 1. INTRODUCTION

Movies business depends on buzz created through promotional activities for that motion picture. Different handles like TV shows, social media and blog spots can be used for promotional activities. There are many users sharing their opinions and experiences via social media, there is an aggregation of personal wisdom and different viewpoints. Such aggregation has limitations as viewpoints and are subject to change with time. In a sense the social media prediction problem is paralleled by prediction of financial time series based on past history, which has its uses in trading. In this project, machine learning algorithms are used for predictive analysis. Machine learning algorithms applied on conventional data, collected from movies databases, and social media features (text comments, tweets, etc). In general, if extracted and analyzed properly, the data on social media can lead to useful predictions of certain human related events. Such prediction has great benefits in many realms, such as finance, product marketing and politics, which has attracted increasing number of researchers to this subject. Study of social media also provides insights on social dynamics and public health. A

survey provides us perspective and is helpful for carrying out further research. Prediction of success in motion picture business has been of great interest.

In a sense the social media forecasting problem is complemented by prediction of financial time series based on past history, which has its uses in trading business. Thus social media plays a vital role in predicting the success of a motion picture.

## 2. LITERATURE SURVEY

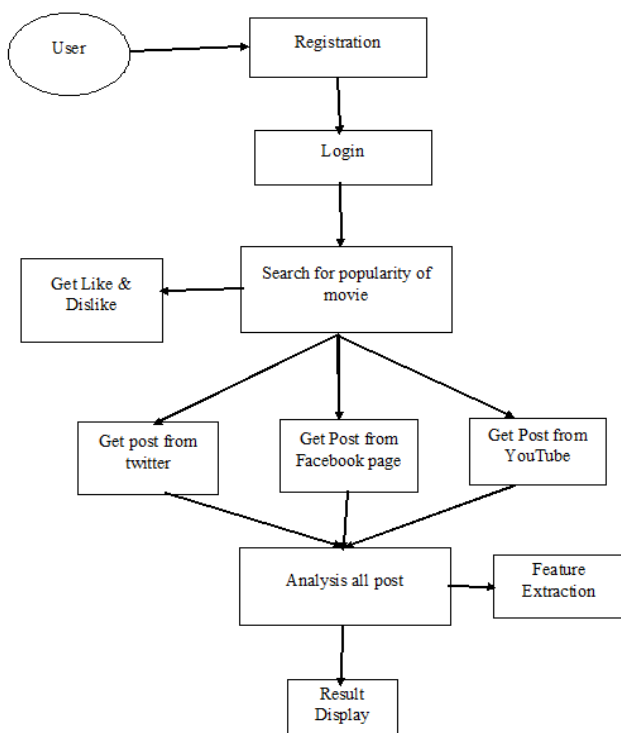
In a study [1], the author described the classification of movies from flop to super doper hit over the years. They have used neural network algorithm on as many as 8 independent variables and found that number of screens, high technical effects and high value of star provide a great deal to a movie's success. K-Means clustering, Polynomial and Linear Regression [2] was used on 2510 movies released 1990 henceforth to study and frame a predictive model to get the expected dividend. They achieved efficiency of 36.9%.

Another study tested Text regression on critics' film reviews to foresee the initial weekend earnings for them metadata collected for 2005-2009 movies. The dataset included of 1718 motion pictures [3]. The authors adopted 7 metadata features including Movie Running Time (in minutes), cost of production(budget), the number of initial weekend screens, genre, MPAA rating, opening occasion (whether summer or holiday), total number of artists, high grossing actors count and if the motion picture had any Oscar winning artists and directors. The same way, 3 types of text features were obtained from the meta-data features.

The authors[10] put forth the concept to mix classical and social media parameters to enhance the prediction efficiency of the movie failure/success. They gathered social characteristics (Posts, Tweets, views) from social networking sites like YouTube, Twitter, Facebook. The study gives out that by increasing the data set, a higher efficiency than the one procured (70%) via linear regression, can be obtained. They collected the id, tweets text, time, username and method from Twitter API and also searched for the matching movie tweets. Ling pipe sentiment analyzer was made use for Sentiment analysis on the tweets, to classify movies as flop, average and super hit.

An efficiency of 64.4% was obtained. The twitter platform sometimes has noisy data and the analyzer used may not be suitable for this. Multiple experiments were done to check whether we could predict the after release movie collections successfully through Twitter (Asur, Sitaram, and Bernardo Huberman "Predicting the future with social media." Web Intelligence and Intelligent Agent Technology (WIIAT), 2010 IEEE/WIC/ACM International Conference on. Vol. 1. IEEE, 2010). Sitaram and Bernardo extracted 2.89 million tweets from 24 movies in 2009 using the twitter's API. The conclusion was that the effect of tweets done for promotion was negligible while the tweet mentions per hour for the motion picture predicted precise box office results. After predicting first weekend revenue they calculated the subjectivity and polarity of the movies by applying sentiment analysis on the tweets. Although the sentiments did improve the results, they were not as important as the tweet rate.

### 3. PROPOSED SYSTEM FLOW



### I. Neuro Linguistic programming (NLP)

Created by Richard Bundler and John Grinder in California, United States in the 1970s, Neuro Linguistic Programming (NLP) is an approach to communication, personal development, and psychotherapy. NLP has since been overwhelmingly discredited scientifically, but continues to be marketed by some hypnotherapists and by some companies that organize seminars and workshops on

management training for businesses. NLP's creators claim there is a connection between neurological processes (neuro-), language (linguistic) and behavioral patterns learned through experience (programming), and that these can be changed to achieve specific goals in life. Bandler and Grinder also claim that NLP methodology can "model" the skills of exceptional people, allowing anyone to acquire those skills. They claim as well that, often in a single session, NLP can treat problems such as phobias, depression, tic disorders, psychosomatic illnesses, near-sightedness, allergy, common cold, and learning disorders. There is no scientific evidence supporting the claims made by NLP advocates and it has been discredited as a pseudoscience by experts. Scientific reviews state that NLP is based on outdated metaphors of how the brain works that are inconsistent with current neurological theory and contain numerous factual errors. Reviews also found that all of the supportive research on NLP contained significant methodological flaws and that there were three times as many studies of a much higher quality that failed to reproduce the "extraordinary claims" made by Bandler, Grinder, and other NLP practitioners. Even so, NLP has been adopted by some hypnotherapists and also by companies that run seminars marketed as leadership training to businesses and government agencies.

### Application Programming Interface: Twitter

The Twitter's API is a set of web addresses that takes factors. These addresses or URLs let you access many features of Twitter, such as posting a tweet or finding tweets that contain a word, etc. The API is documented here, the platform is a social media network and communication application that gives more than 200 million tweets a day. The Twitter platform offers access to that enormous amount of data. All the API's represent a facet of Twitter, and give developers a freedom to develop and extend their applications in a new and creative ways. The Twitter APIs are constantly being worked upon, and creating something on this Platform is not a one-off event. Twitter sets its application programming interface (API) off the Representational State Transfer (REST) structure. This architecture refers to a collection of network design principles that define resources and ways to address and access data. It is a design philosophy, not just blueprints, there's variable arrangement of computers, servers and cables. For Twitter, a REST architecture in part means that the service works with most Web syndication formats.

### Application Programming Interface: Facebook

The Facebook API is the Graph API that allows user to read and write data from Facebook. Facebook also features the Old Rest API. The upcoming Graph API converts the Application Programming Interface paradigm from a method-oriented way of reading and writing data to a new way that uses objects (think profiles, FB friends, FB posts, shared photos, comments & likes, and so on) and their connections with one another. This approach makes the FB

API easy and more agreeable when working with objects. The Old REST API is active although the Graph API is the preferred. Both the APIs are relevant to mobile applications. Graph API objects are assigned a unique ID and are easily addressable using a URL that can be further qualified to address specific of object/connection. Graph.facebook.com/OBJECT\_ID/CONNECTION\_TYPE Where OBJECT\_ID is the object's unique ID and CONNECTION\_TYPE is one of the connection types supported by the object. For example, a FB page provides the following data: Newsfeed, shared photos, liked posts, friends, etc. Graph API helps you to receive an object, delete it, and publish the objects. You can do operations such as search and update objects, filter out the results, and even find out the connections of an object by default. In order to obtain restricted/private data, applications must first ask the user's consent, specified as extended permissions. One can find these consent details on Extended Permissions page.

## II. TECHNOLOGIES TO BE USED

### JAVA

Java is a programming language expressly designed for use in the distributed environment of the Internet. It was designed to have the "look and feel" of the C++ language, but it is simpler to use than C++ and enforces an object-oriented programming model. Java can be used to create complete applications that may run on a single computer or be distributed among servers and clients in a network. It can also be used to build a small application module or applet for use as part of a Web page. Applets make it possible for a Web page user to interact with the page. In the early nineties, Java was created by a team led by James Gosling for Sun Microsystems. It was originally designed for use on digital mobile devices, such as cell phones. However, when Java 1.0 was released to the public in 1996, its main focus had shifted to use on the Internet. It provided more interactivity with users by giving developers a way to produce animated webpages. Over the years it has evolved as a successful language for use both on and off the Internet.

### Tomcat

Apache Tomcat, or generally known as Tomcat, is an open-source web server. It was developed by the Apache Software Foundation. Tomcat enforces various Java EE modules such as Java Servlet, Java Server Pages, Java Expression Language, and Web Socket, and helps us with a "Java Code" HTTP web server environment in which the code executes. An open community of developers maintain and develop Tomcat under the protection of the ASF, published under the Apache License 2.0. It is an open-source software. Catalina is Tomcat's servlet container. Catalina implements Sun Microsystems' specifications for servlet and Java Server Pages (JSP). In Tomcat, an element represents a database of usernames, passwords, and roles (similar to UNIX groups) assigned to those users. Different implementations of these elements allow Catalina to be integrated into environments

where such authentication data is already being made and curated, and then use that data to implement Container Managed Security as described in the Servlet Specification. It has also added user- as well as system-based web applications enhancement to add support for deployment across the variety of environments. It also tries to manage sessions as well as applications across the network.

### Hibernate

Hibernate ORM is object-relational mapping framework for the Java. It provides a framework for mapping an object-oriented domain model to a relational database. Hibernate solves object-relational impedance mismatch problems by replacing direct, persistent database accesses with high-level object handling functions.

Hibernates primary feature is mapping from Java classes to database tables; and mapping from Java data types to SQL data types. Hibernate also provides data query and retrieval facilities. It generates SQL calls and relieves the developer from manual handling and object conversion of the result set.

## 4. CONCLUSION

This system represents the comparison of Conventional Features with Social Media features in determining the popularity of movies. The social media features such as Sentiment Score of tweets related to movies, Number of Views and Comments of movies' trailer on YouTube and fan following on twitter can usefully be utilized to predict the popularity of movie. We can fetch that comments or posts from social media and perform a semantic analysis to see if movie making positive hype or not. We can use feature extraction to get these details and predict popularity of movie.

## 5. FUTURE SCOPE

In the future we can hope to build a system that can give us the most accurate prediction and review of movies using more than one parameters and eliminate the need of multiple websites.

## REFERENCES

- [1] Sharda, R., & Delen, D. (2006): "Predicting box-office success of motion pictures with neural networks". *Expert Systems with Applications*, 30(2), 243-254.277.
- [2] Asur, Sitaram, and Bernardo Huberman. "Predicting the future with social media." *Web Intelligence and Intelligent Agent Technology (WIIAT)*, 2010 IEEE/WIC/ACM International Conference on. Vol. 1. IEEE, 2010.
- [3] Apala, Krushikanth R., et al. "Prediction of movies box office performance using social media." *Advances in Social Networks Analysis and Mining (ASONAM)*, 2013 IEEE/ACM International Conference on. IEEE, 2013. 278

[4] Fang Wu, Dennis Wilkinson and Bernardo A. Huberman. Feedback Loops of Attention in Peer Production. Proceedings of SocialCom-09: The 2009 International Conference on Social Computing, 2009.

[5] "Predicting Movie Trailer Viewer's "Like/Dislike" via Learned Shot Editing Patterns" Yimin Hou, Ting Xiao, Shu Zhang, Xi Jiang, Xiang Li, Xintao Hu, Junwei Han, Lei Guo, L. Stephen Miller, Richard Neupert, and Tianming Liu IEEE Transactions on affective computing 2016.

[6] Mehreen Ahmed , Maham Jahangir , Dr. Hammad Afzal "Using Crowd-source based features from social media and Conventional features to predict the movies popularity" 2015 IEEE International Conference

[7] Rui, Huaxia, Yizao Liu, and Andrew Whinston. "Whose and what chatter matters? The effect of tweets on movie sales." Decision Support Systems 55.4 (2013): 863-870.

[8] Bhave, A., Kulkarni, H., Biramane, V., & Kosamkar, P. (2015). "Role of different factors in predicting movie success". In Pervasive Computing (ICPC), 2015 International Conference on (pp. 1-4). IEEE.