

Movies Popularity Prediction Using Social Media and Conventional Features

Babita M. Jangid¹, Chaitali K.Jadhav¹, Swati M. Dhokate¹, Grish M.Jadhav¹, Dr. G.M.Bhandari²

B.E. Students at Department of Computer Engineering, JSPM'S BSIOTR, Wagholi, Pune,
Maharashtra, India¹

H. O. D. Department of Computer Engineering, JSPM'S BSIOTR, Wagholi, Pune, Maharashtra, India²

ABSTRACT: Now a day, use of social media has been increased widely. By using social media like Twitter, YouTube etc. users can posts their reviews about movies. The economists, investors and predictive analysts are very interested to predict the success of their movies. For predicting success as well as popularity of movies number of factors affected like actor, actress, invested budget, production house, genre, PG rating. In this project, machine learning algorithms are used for predictive analysis. Machine learning algorithms applied on conventional, collected from movies databases, and social media features (text comments on Tweets, YouTube). Mining the attributes and contents of social media gives us an opportunity to discover social structure characteristics, analyse action patterns qualitatively and quantitatively, and sometimes the ability to predict future human related events. Result of this project that predicts the success with control and use of sentiments form social media and other social media features.

KEYWORDS: Data Mining, Predictive Analysis, Classification, Regression, social media, social network, social networking service, user-generated contents.

I. INTRODUCTION

Parallel There are many users sharing their opinions and experiences via social media, there is aggregation of personal wisdom and different viewpoints. Such aggregation has limitations as viewpoints 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 general, if extracted and analysed 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 business has been of great interest. Social media comprises interactive applications and platforms .Past few years brought huge growth in social media. It aggregates opinions and feelings of diverse groups of people. Ability to predict future human related events. Prediction of success in business has been of great interest. Prediction of success in business has been of great interest to the economists and financial experts. With advent of data analytics, the prediction process has been made intelligent by considering the historical data and employing various data analytical techniques to infer the future events. Such studies have been performed in prediction of movies success as well where success and popularity is measured in terms of the Ratings (typically represented by a numeric number from 0-10) and Income.

We found that most discriminating feature that played vital role in classifying the Ratings and Income was Sentiment score. One of our proposed feature, Aggregate Actor Followers, performed better than traditionally used top Actor followers. Other features proposed including Likes/Dislikes on YouTube, Sequel also played vital role in prediction .We used these attributes to measure

- (i) Rating of movie,
- (ii) Box Office success of movie.

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 4, April 2017

II. RELATED WORK

In [1] author was conducted over a span of five years (1998-2002) in which the authors classified nine classes from flop to blockbuster. They applied neural network algorithm on 7 independent variables and found that number of screens, high technical effects and high star value contribute a great deal to a movie's success. K-Means clustering, Polynomial and Linear Regression [2] was applied on 2510 movies released 1990 onwards to study and build a predictive model to get the expected revenue. They achieved accuracy of 36.9%. Another study [3] applied Text regression on critics' film reviews to predict the opening weekend revenue for the metadata collected for 2005-2009 movies. The dataset consisted of 1718 movies. The authors used seven metadata features including Movie Running Time (in minutes), Budget, the number of opening weekend screens, genre, MPAA rating, opening time (whether summer or holiday), total number of actors, high grossing actors count and whether the movie had any Oscar winning actors and directors. Similarly three types of text features were extracted from the metadata features. For the first weekend release revenue metadata features gave an accuracy of 0.521 and the amalgamation of text, metadata features gave even better results. In [4] researchers proposed the idea to integrate classical and social media factors to improve the prediction accuracy of the movie success. They collected classical attributes (genre, budget etc.) from IMDB and social attributes (Tweets, views) from social websites like YouTube, Twitter. The study suggests that by increasing the data set, a higher accuracy than the one obtained (70%) through linear regression, can be achieved. In [5], authors predicted the first weekend box office revenue for movies released in 2010. They used a data set of 312 movies collected from Box Office Mojo and the attributes including views count, editors' count, number of edits and collaborative rigor from Wikipedia articles. The opening weekend revenue and number of theatres screens were also included. They applied linear regression and got an accuracy of 0.94 one month prior to release date of the movie. In [6] author used an existing data set of 2009, 2012 movies provided by SNAP, a Stanford university research group. They collected the tweets text, id, username, time and method from Twitter API and searched for the relevant movie tweets. Ling pipe sentiment analyzer was used for Sentiment analysis on the tweets, to classify movies as hit, flop and average. An accuracy of 64.4% was computed as tweets can have noisy data and the analyzer used was not suitable for tweets. Author [7] apply data mining tools to generate interesting patterns for predicting box office performance of movies using data collected from multiple social media and web sources including Twitter, YouTube and the IMDB movie database. The prediction is based on decision factors derived from a historical movie database, followers count from Twitter, and sentiment analysis of YouTube viewers comments. Social broadcasting networks such as Twitter in the U.S. and "Weibo" in China [9] are transforming the way online word of mouth (WOM) is disseminated and consumed in the digital age. This paper investigated whether and how Twitter WOM affects movie sales by estimating a dynamic panel data model using publicly available data and well-known machine learning algorithms.

III. PROPOSED SYSTEM

In proposed system, the use of social media as well as conventional features like genre, ratings, gross, sequel of movie, budget, screens to predict the movie popularity as well as revenue. Social media like Twitter and YouTube are used as database to proposed system. Proposed system comprises main two stages, namely data collector and Predictive engine.

1. Data Collector:

Data collector is the main module in proposed system. It collects the information about the movies from YouTube and Twitter also from web sites i.e. IMDB, generic web resource i.e. Wikipedia.

2. Feature Extraction:

This is second major module of proposed system that extracts the features by using sentiments. It extracts the features from social media and conventional features. This module represents sentiment score of a movie of tweets which are linked to respective movie. This module also finds the polarity of the tweets.

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 4, April 2017

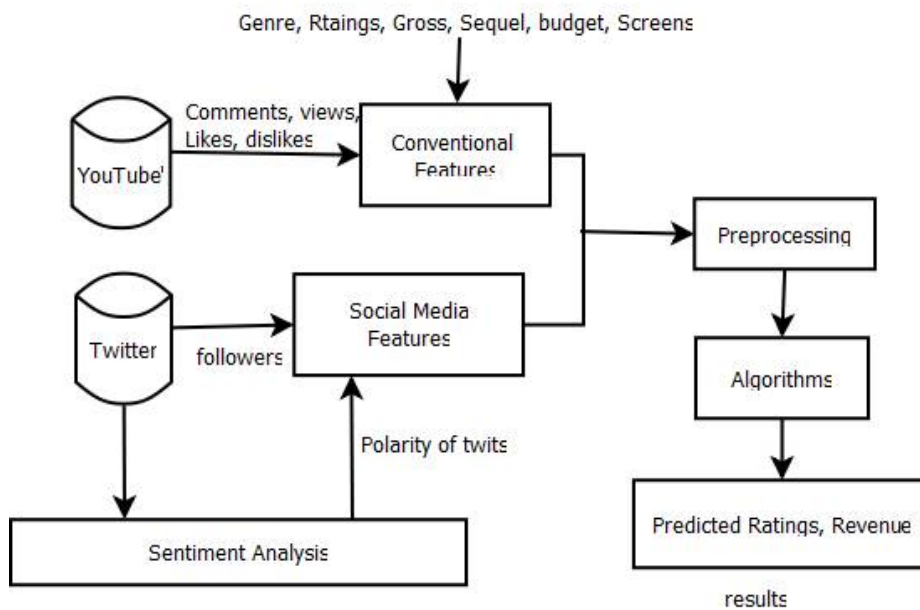


Fig1: System architecture of proposed system.

3. Conversions:

This module dealing with maximum values, data transformation (converting currencies, sequel to 1-N, genre to [1...19] etc.) which is very useful in calculation sentiment analysis score for each tweet etc.

4. Prediction:

In this module, system can predict the ratings, bands of ratings, bands of popularity, income of the movies. Here numbers of algorithms are used like linear, polynomial regression, ID3, KNN.

IV. EXPERIMENTAL RESULTS

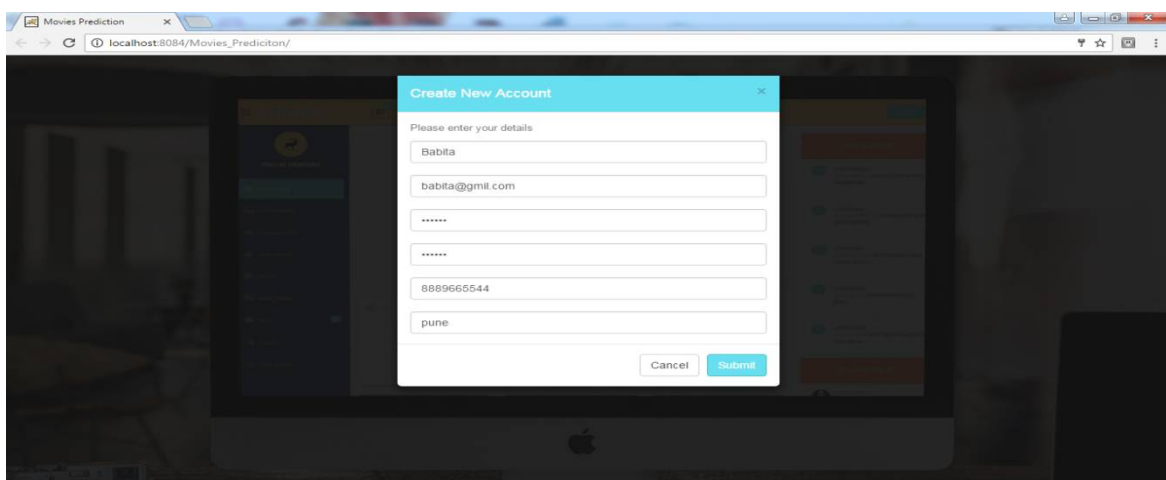


Fig2: Registration page

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 4, April 2017

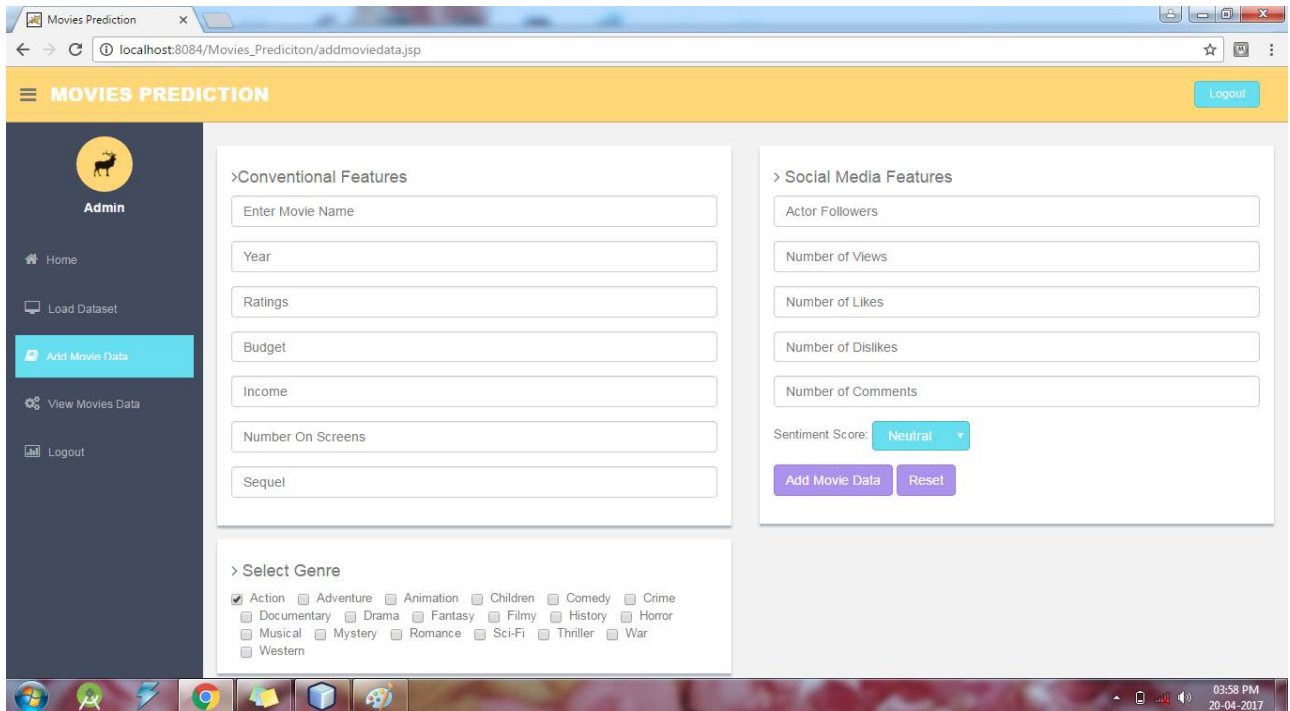


Fig3: Add movie details

In this figure show add movie details like movie name, release date/year, rating, budget, how much state release? All of the information adds this figure.



Sr No	Movie Name	Year	Genre	Rating	Budget	Income	Screens	Sequel	Followers	Views	Likes	Dislikes	Comments	Score
1	Pirates of the Caribbean: At Worlds End	2007	Action, Adventure, Fantasy	7.1	300000000	309404152	11500	2	40000	586143	1839	29	295	positive
2	Spider Man 3	2007	Action, Adventure, Romance	6.2	258000000	336530303	1030	3	24000	3362412	3933	465	2260	negative
3	Harry Potter and the Half Blood Prince	2009	Adventure, Family, Fantasy, Mystery	7.5	250000000	301956980	12000	3	25000	1382675	4351	113	2500	positive
4	Batman v Superman Dawn of Justice	2016	Action, Adventure, Sci-Fi	6.9	250000000	330249062	30000	4	15000	68279136	432927	17024	98932	neutral
5	Superman Returns	2006	Action, Adventure, Sci-Fi	6.1	209000000	200069408	11000	4	18000	872270	2232	192	935	negative
6	Man of Steel	2013	Action, Adventure, Fantasy, Sci-Fi	7.2	225000000	291021565	13200	3	15000	40780339	145669	6178	64983	neutral
7	The Avengers	2012	Action, Adventure, Sci-Fi	8.1	220000000	623279547	11800	1	26000	25768212	84998	2489	31650	negative
8	Pirates of the Caribbean: On Stranger Tides	2011	Action, Adventure, Fantasy	6.7	250000000	241063875	18210	4	40000	2962970	8256	174	2493	positive
9	Men in Black 3	2012	Action, Adventure, Comedy, Family, Fantasy, Sci-Fi	6.8	225000000	179020854	10920	3	10000	1014320	2366	92	514	negative
10	Robin Hood	2010	Action, Adventure, Drama, History	6.7	200000000	105219735	4000	1	891	290936	197	18	25	positive
11	King Kong	2005	Action, Adventure,	7.2	207000000	218051260	5032	1	6000	1053441	1798	132	450	neutral

Fig4: Rating of movies

International Journal of Innovative Research in Science, Engineering and Technology

(An ISO 3297: 2007 Certified Organization)

Website: www.ijirset.com

Vol. 6, Issue 4, April 2017

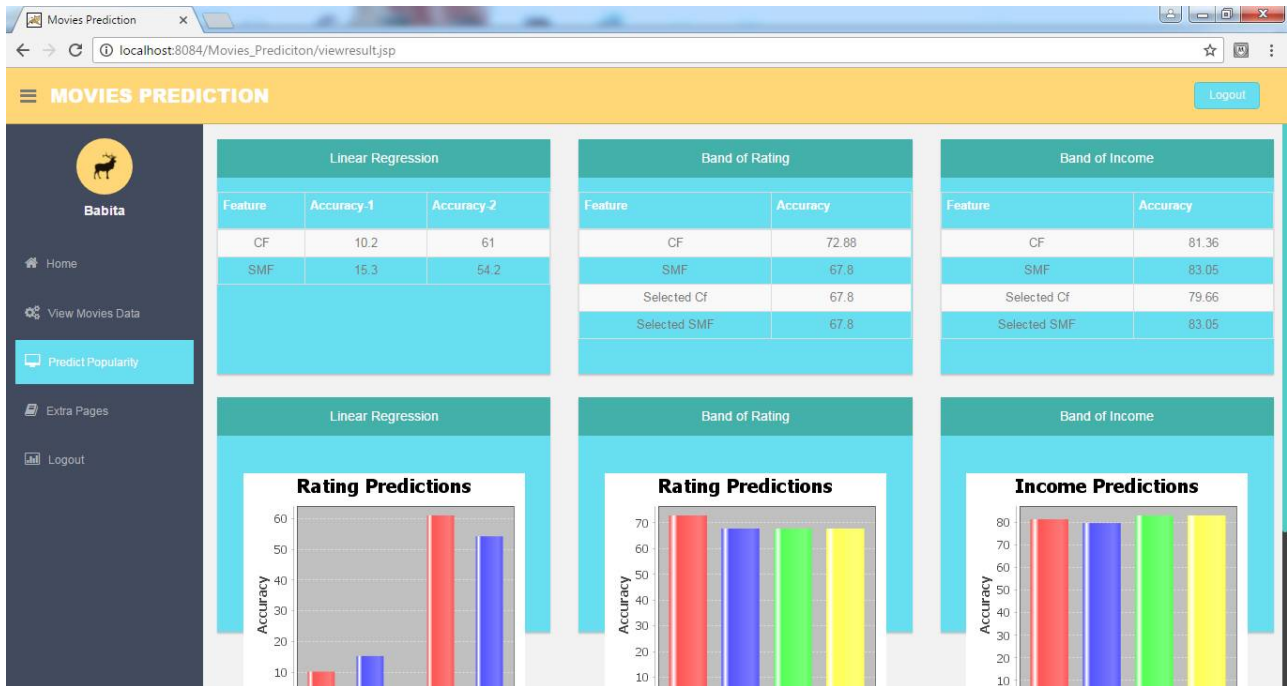


Fig5.Linear regression-rating prediction.

V. CONCLUSION

A promising step towards inferring the impacts of social media on crowd funding. We anticipate that further analysis could quantify personal interests of participants from historical behaviors, develop richer ways of assessing social influences among users, and suggest potential customers from collaborative filtering of personal interests and social influences. The comparison of Conventional Features with Social Media features in determining the popularity of movies. Social media features such as Sentiment Score of tweets related to movies, Number of Views and Comments of movies trailers on YouTube and fan following on twitter can usefully be utilized to predict the popularity of movie. However, most of these studies were performed either on small set of data (less than 40 movies), or they considered tweets of only 1 week/1 month before release. Our experiments showed that, on larger data sets, with no consideration of time while collecting statistics from twitter and YouTube, social media features give better performance.

REFERENCES

- [1] Nikhil Apte, Mats Forssell, and A. Sidhwa, "Predicting Movie Revenue". 2011.
- [2] Joshi, M., Das, D., Gimpel, K., & Smith, N. A. (2010). "Movie reviews and revenues: An experiment in text regression". In Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics (pp. 293-296). Association for Computational Linguistics.
- [3] Bhawe, 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.
- [4] Mestyán, Márton, Taha Yasseri, and János Kertész. "Early prediction of movie box office success based on Wikipedia activity big data." PLoS one 8,8 (2013): e71226.
- [5] Jain, Vasu. "Prediction of Movie Success using Sentiment Analysis of Tweets." The International Journal of Soft Computing and Software Engineering 3.3 (2013): 308-313.
- [6] Asur, Sitaram, and Bernardo Huberman. "Predicting the future with social media." Web Intelligence and Intelligent Agent Technology (WI-IAT), 2010 IEEE/WIC/ACM International Conference on. Vol. 1. IEEE, 2010.
- [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] 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.
- [9] Sharda, R., & Delen, D. (2006). "Predicting box-office success of motion pictures with neural networks". Expert Systems with Applications, 30(2), 243-254.