

## Chapter 8

# Prediction of Movie Success Using Sentimental Analysis and Data Mining

**Meenu Vijarania**

*Center of Excellence, K. R. Mangalam University, Gurugram, India*

**Ashima Gambhir**

*Amity University, Gurgaon, India*

**Deepthi Sehrawat**

*Amity University, Gurgaon, India*

**Swati Gupta**

*Center of Excellence, K. R. Mangalam University, Gurugram, India*

### ABSTRACT

*Movies have become a significant part of today's generation. In this chapter, the authors worked on data mining and ML techniques like random forest regression, decision tree regression, support vector regression, and predict the success of the movies on the basis of ratings from IMDb and data retrieved from comments on social media platforms. Based on ML techniques, the chapter develops a model that will predict movie success before the release of the movie and thereby decrease the risk. Twitter sentimental analysis is used to retrieve data from Twitter, and polarity and subjectivity of the movie is calculated based on the user reviews, and those retrieved data machine learning algorithms are used to predict the IMDb rating. A predictive model is developed by using three algorithms, decision tree regression, SVR, and random forest regression. The chapter compared the results using three different techniques to get the movie success prediction at a reasonable accuracy.*

DOI: 10.4018/978-1-7998-9012-6.ch008

## **INTRODUCTION**

Movies are one of the convenient ways of one's entertainment. Irrespective of age love for movies remains the same, whether he is an older man or a schoolboy. Movie industry produces a lot of movies every year, but a few become a success, few gain medium exposure while rest come under the category of a flop. A lot of investment takes place in completing and finalizing a movie having many star casts and many workers working day and night for the success of it. The revenue that will be generated by the movie after it's release depends on many factors like casts of the movie, budget of the movie, review given by the critics, rating of the movie, the year movie released, time of the release etc. The parameters involved in predicting the movie success made it difficult to find a direct formula for the prediction of the success. However, a model can be built by choosing those parameters which can directly affect the forecast, which will be beneficial for the businessmen also and the viewers even. Based on the past experiences and reviews of similar movies and casts, it can be predicted whether a movie will be a success or a flop. It will help the stakeholders to invest correctly and will be beneficial for the audiences also in selecting a movie. They may be able to make the decision before the release of the movie.

Based on machine learning techniques, the proposed work develops a model which will help in the prediction of movie success before the release of the movie and thereby decreasing the risk somehow. The best way to know about movies and its revenue generation is IMDb ratings. In this model, Sentimental Analysis is done on the Twitter data, and the relation between polarity and subjectivity is calculated. On that calculated information, three Machine learning Algorithms are applied and compared to get the best accuracy in prediction of the IMDb ratings. The dataset has been taken from Kaggle.com Vast amount of data is present there about the movies, there gross collection, critics reviews, working cast members and factors influencing their work. Machine learning techniques enabled us to discover many aspects which is helpful in the prediction of success or failure of upcoming movies based on some information provided beforehand. Here we tried to develop a model which will predict movie success rate with more accuracy and precision. This model will predict the IMDb rating of Hollywood as well as Bollywood movies. Python programming language is used for making this model. We divide the dataset into a training set and testing set for the same. From the result we get, we can easily conclude that the movie will be successful or not. There is a twofold outcome of this research, firstly providing tools and techniques which can transform data into a suitable format for machine learning, secondly, provides selected information taken from this refined data.

14 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: [www.igi-global.com/chapter/prediction-of-movie-success-using-sentimental-analysis-and-data-mining/302066](http://www.igi-global.com/chapter/prediction-of-movie-success-using-sentimental-analysis-and-data-mining/302066)

## Related Content

---

### A Beam Search Based Decision Tree Induction Algorithm

Márcio Porto Basgalupp, Rodrigo Coelho Barros, André C. P. L. F. de Carvalho and Alex A. Freitas (2012). *Machine Learning Algorithms for Problem Solving in Computational Applications: Intelligent Techniques* (pp. 357-370).

[www.irma-international.org/chapter/beam-search-based-decision-tree/67712](http://www.irma-international.org/chapter/beam-search-based-decision-tree/67712)

### Application of Structural Properties of Seismic Data to Prediction of Hydrocarbon Distribution

Zhou Wei, Guo Haimin and Yaoting Lin (2018). *International Journal of Software Science and Computational Intelligence* (pp. 41-52).

[www.irma-international.org/article/application-of-structural-properties-of-seismic-data-to-prediction-of-hydrocarbon-distribution/207744](http://www.irma-international.org/article/application-of-structural-properties-of-seismic-data-to-prediction-of-hydrocarbon-distribution/207744)

### Fuzzy Image Segmentation for Mass Detection in Digital Mammography: Recent Advances and Techniques

Hajar Mohammedsaleh H. Alharbi, Paul Kwan, Ashoka Jayawardena and A. S. M. Sajeev (2012). *Multidisciplinary Computational Intelligence Techniques: Applications in Business, Engineering, and Medicine* (pp. 378-402).

[www.irma-international.org/chapter/fuzzy-image-segmentation-mass-detection/67302](http://www.irma-international.org/chapter/fuzzy-image-segmentation-mass-detection/67302)

### GBF Trained Neuro-Fuzzy Equalizer for Time Varying Channels

Archana Sarangi, Sasmita Kumari Padhy, Siba Prasada Panigrahi and Shubhendu Kumar Sarangi (2013). *Modeling Applications and Theoretical Innovations in Interdisciplinary Evolutionary Computation* (pp. 176-188).

[www.irma-international.org/chapter/gbf-trained-neuro-fuzzy-equalizer/74929](http://www.irma-international.org/chapter/gbf-trained-neuro-fuzzy-equalizer/74929)

### Methods of Information Processing of Relative Motion in the Flying Groups of UAV

Rinat Galiautdinov (2021). *International Journal of Software Science and Computational Intelligence* (pp. 90-108).

[www.irma-international.org/article/methods-of-information-processing-of-relative-motion-in-the-flying-groups-of-uav/266230](http://www.irma-international.org/article/methods-of-information-processing-of-relative-motion-in-the-flying-groups-of-uav/266230)