MOVIE SUCCESS PREDICTION USING DATA MINING

[1]Hrithik Jain, [2]Manohar Bhati, [3]Sailee Zodape [4]Blessy Varghese

[123]Student, [4]Professor

[123] Department of Information Technology, [4] Department of Computer Engineering

[1]Pillai College of Engineering University of Mumbai, Navi Mumbai, India

Abstract: Movies are released in great numbers every year and a huge amount of capital is invested to produce those films. The main focus of our project is to predict whether the movie will be successful, i.e. hit or unsuccessful, i.e. flop, based on certain attributes (static and social/dynamic). Few of the static factors included for determining whether the movie will be successful or not are: genre, film budget, actors, actresses, director, producer, production house, release date etc. From the social media perspective, i.e. dynamic, we would be looking at trending hashtags on Twitter related to the movie. As a huge amount is invested in making a film, it is a concerning aspect for the production house, and thus would be beneficial for the investors if a certain prediction is given. The primary aim is to find the relation between the attributes to arrive at the conclusion using a data mining algorithm. A movie score is calculated by comparing a movie's attributes with other movies in a dataset with similar attributes that provide a good deal of prediction. Various mechanisms available in data mining are used to identify whether each film is good or bad. This system is especially useful for producers because it allows them to rate movies before they are released and change the way they post-produce to increase their chances of success**.**

Index Terms - **Movie Success Prediction, Data Mining, Twitter.**

I.INTRODUCTION

Movie success prediction will help producers understand that the investment made by them in the movie is going to be worth it or not. Data mining and machine learning algorithms will analyze all the past data of the components and by using various machine learning algorithms the model can predict the result of the movie. Past performance, budget, date of release, actors, actresses and director will contribute towards predictions by training the model on the basis of the past data. Adding dynamic factors to be more precise about the prediction, twitter tweets and hashtags are analyzed on the basis of sentiments. This sentiment analysis returns positive, negative and neutral hype as results for the movie.

II. LITERATURE SURVEY

2.1 Movie Success Prediction Using Data Mining

This project focuses on devising and developing a mathematical model to predict the performance of movies. The mathematical approach for prediction of successful movies is the major idea of the project.

2.2 Data mining Technique for Analyzing and Predicting the success of Movie

An attempt is made to predict the past as well as the future of movie for the purpose of business certainty or simply a theoretical condition in which decision making [the success of the movie] is without risk, because the decision maker [movie makers and stakeholders] has all the information about the exact outcome of the decision, before he or she makes the decision release of the movie.

2.3 Box-Office Opening Prediction of Movies based on Hype Analysis through Data Mining

Pre-Release hype is an important factor for estimating the openings of the movie. This can be estimated through user opinions expressed online on sites such as Twitter which is an online micro-blogging site with a user base running into millions.

III. PROPOSED WORK

The proposed system aims towards creating a robust and more reliable predictor by using the maximum amount of data possible. Existing systems only used static parameters to predict the success of a movie. Considering only a dataset with past data and training the model according to that data to provide an output is one effective way to perform the operation. The proposed model aims to include both static and dynamic factors to improvise results. Dynamic factors include sentiment analysis technology. Using twitter hashtags and tweets as data related to the keyword i.e the name of the movie and analyzing the tweets to project out positive, negative or neutral sentiment about the movie. This factor is useful to improve the scope of prediction as its real time data and not past dataset.
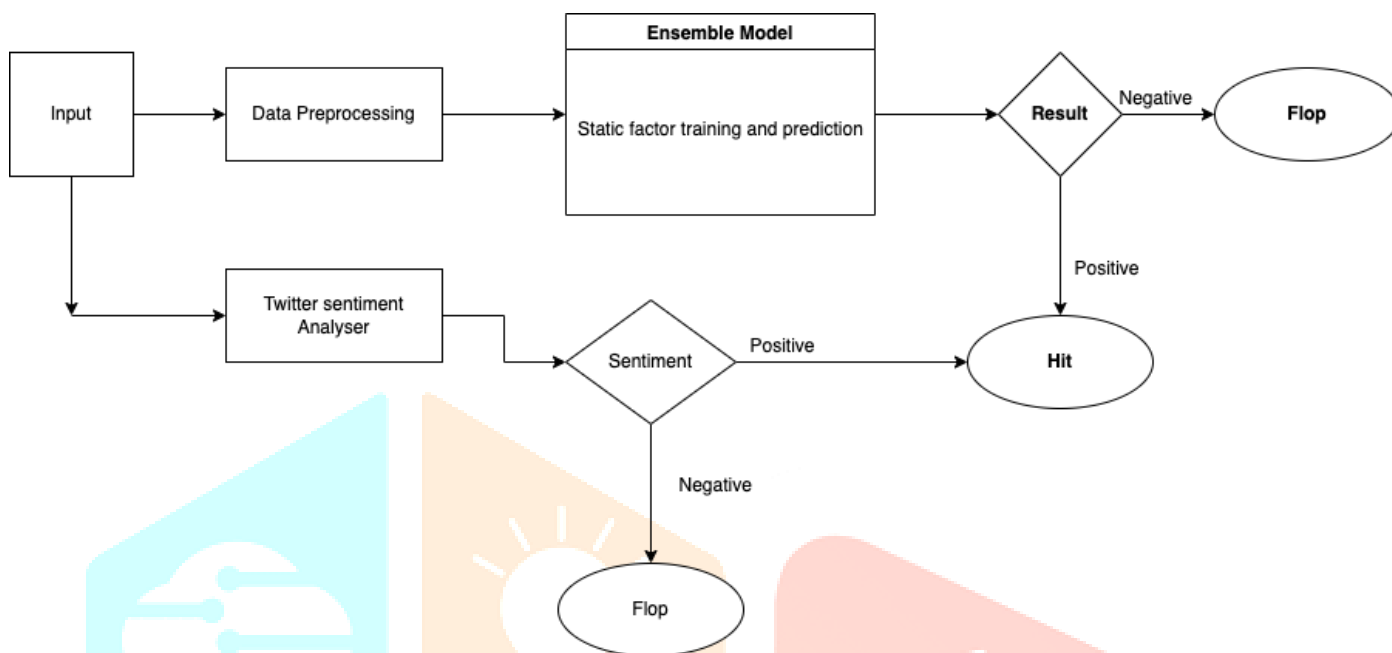

Fig 1. Pictorial representation of Movie Success Prediction

IV. SYSTEM ARCHITECTURE

1. Ensemble Model
   It has two models to perform data mining processes based on budget and rating.
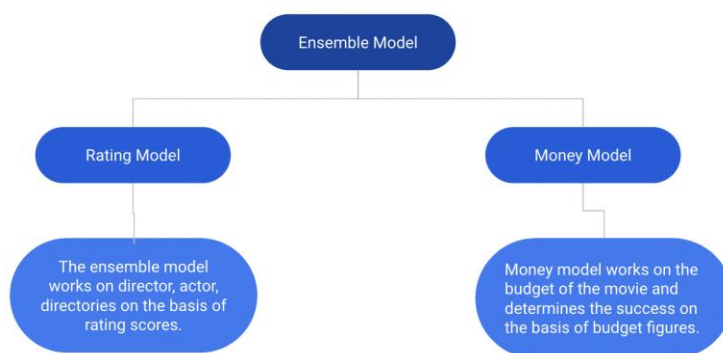   - Rating Model
   - Money Model

Fig 2. Pictorial representation of Ensemble Model

2. Twitter Analyser

Twitter analyzer is responsible for analyzing sentiments of tweets on twitter using hashtags on movie data.
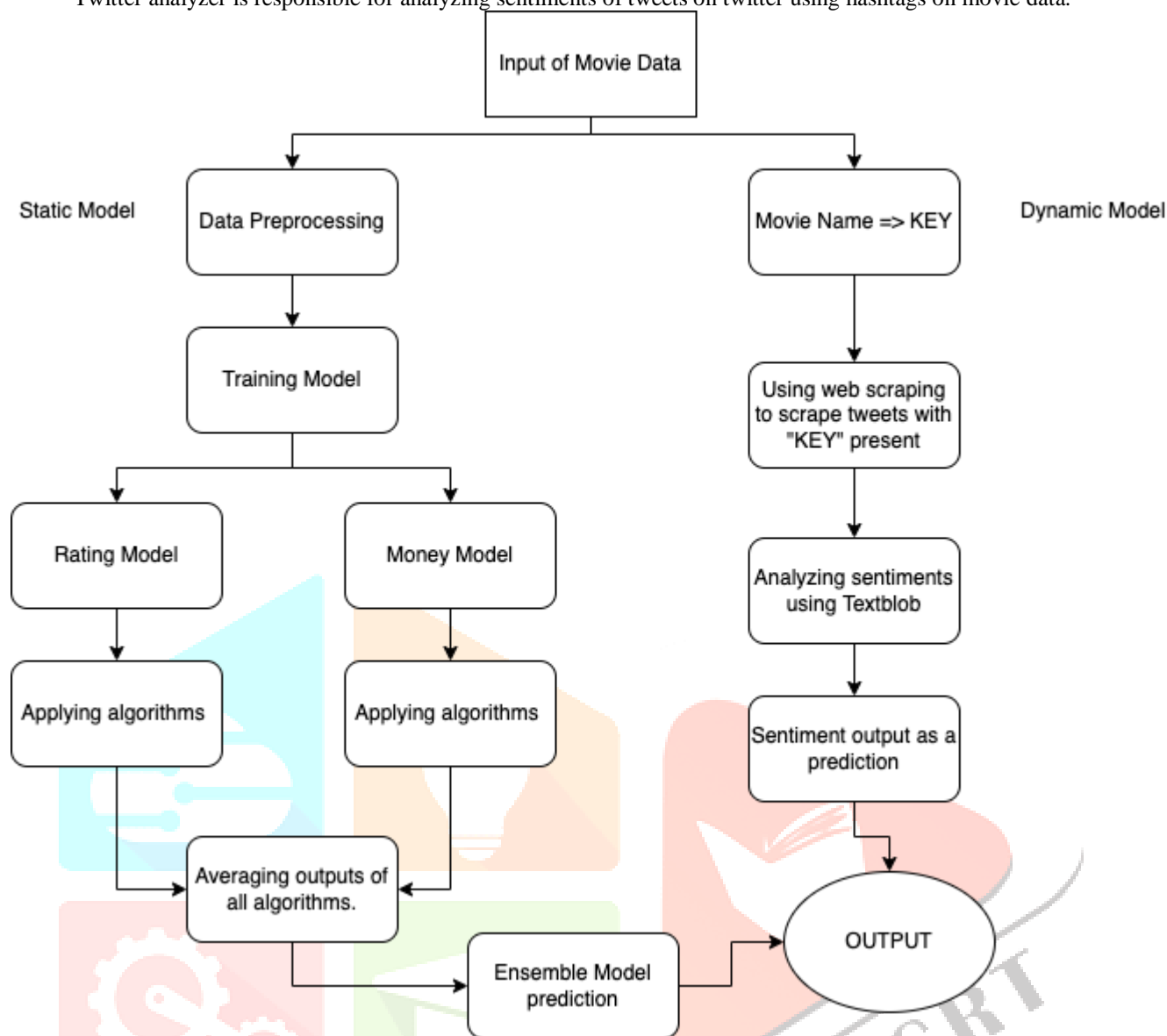

Fig 3. Pictorial representation of system Architecture

a. **Static Block**

Data is preprocessed using Data Cleaning, Data Extraction and Removing Outliers procedures.
Produced dataset goes under algorithms for training and creating predictions under both money and rating models.
Output is averaged out to give precise predictions. This output is ensemble model prediction.

b. **Dynamic Block**

The name of the movie is assigned as a key and then using web scraping tweets are scraped having the movie name or mentioning the key. Then using textblob the fetched statements and sentiments are analyzed to give positive, negative and neutral predictions for hype.

V. RESULT AND CONCLUSION

The system architectures, implementation details ie. algorithms and system requirements, result and scope of the project is presented. Two different system models using the static factors i.e. the rating model and the money model which are used for the prediction of the success of the movie are described. The dynamic components of twitter analysis provide prediction of social media success or failure value. Simple UI and non technical user experience have been achieved for both types of users i.e creator and watcher of the movie. Accuracy of 85% has been observed in the project for predictions of movies yet to be released on the basis of algorithms and twitter sentiment analysis.
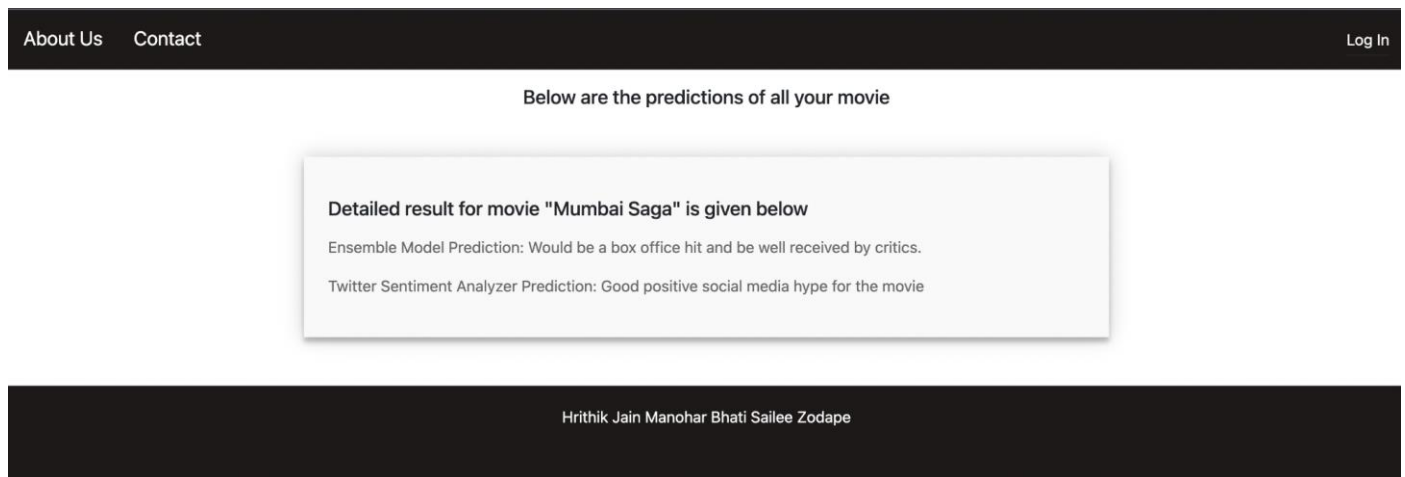

Fig 4. Prediction about the movie

VI. ACKNOWLEDGEMENT

We would like to express our gratitude and thanks to Dr. Sandeep Joshi, Principal, at Pillai's College of Engineering for providing us with a good project guide during a lot of consultation. The help and guidance he gives us from time to time inspiration to complete the project.We also take this opportunity to convey a deep feeling of gratitude to Dr. Satishkumar Varma, HOD Information Technology to the department, with his warm support, valuable information and guidance, which helped to complete the project in various ways categories. Thank you for your cooperation in completing our work. We take this opportunity to extend my deepest gratitude and deepest greetings to our Guide Prof. Blessy Varghese presenting his work ethic and his exemplary guidance, vigilance and ongoing encouragement throughout Of course We would like to thank all the people for their direct and indirect assistance in completing our work.

VII. REFERENCES

[1] Ahmad, Javaria, Prakash Duraisamy, Amr Yousef, and Bill Buckles. "Movie success prediction using data mining." In 2017 8th International Conference on Computing, Communication and Networking Technologies (ICCCNT), pp. 1-4. IEEE, 2017.

[2] Meenakshi, K., G. Maragatham, Neha Agarwal, and Ishitha Ghosh. "A Data Mining Technique for Analyzing and Predicting the Success of a Movie." In Journal of Physics: Conference Series, vol. 1000, no. 1, p. 012100. IOP Publishing, 2018.

[3] Ajay Siva Santosh Reddy Pratik Kasat Abhiyash JainPublished in : 2018 "Prediction of movies box office performance using social media." International Journal of Computer Applications (0975 – 8887) Volume 56– No.1, October 2012.

[4] www.geeksforgeeks.org/data-mining/