Early Predictions Of Movie Success Using Machine Learning

Abstract :

The area of creating predictive models using machine learning has increased in size in recent years. The market for movies is still big with hundreds of new movies created every year. The purpose of this proposal is to investigate whether it is possible to classify movie rating and box office revenue with metadata available before release. This was done by building a classification model with metadata obtained from the internet such as, budget and what actors are involved, etc. This research proposes a way to predict how successful a movie will be prior to its arrival at the box office instead of listening to critics and others on whether a movie will be successful or not. This research provides a quite efficient approach to predict IMDb score on IMDb Movie Dataset. The precision of the predictions can further be increased with a larger data set with more features.

Introduction:

Film industry involves huge investment in terms of money, time, effort and people. So, it is vital to predict success and failure of movie. For this purpose, Internet Movie Database is used that contains categorical and numerical data. Different algorithms are applied in this dataset to calculate IMDb score. IMDb score is nothing but a measure that predicts success and failure of movie accurately before actual release of the movie. In this work, Machine Learning algorithms are used for prediction purpose. Machines learning determines new algorithms from the previous experiments. Nowadays, hundreds of movies are get produced every year. The movie sector is a massive sector for investment but larger the business sectors, more is the complexity, and it is hard to choose how to invest wisely. Furthermore, significant investments come with more significant risks. As film industry is growing too fast, sufficient amount of data is available on the internet. This makes it an exciting as well as interesting field for data analysis. Predicting a movie's box office success is a very complex but essential task to perform. The definition of success of a movie is not fixed. It is relative. Some movies can be called as successful based on their gross income while some movies can snot be. There are many factors responsible in predicting success of movies. Instead of predicting only hit or flop movies, we can classify a film based on the profit it is making on box office into one of five categories such as super flop, flop, average, hit, super hit. Among them, there exist both good movies and flop ones. Therefore, the main question is how to know quality of movie before actually watching it or how can we choose a best movie to watch and relax on our weekends? Most of the time, we will check the movie score or have a look at its review to make right choice. IMDb website is really a good choice to refer at this time. Due to its popularity, IMDb website provides a great deal of information about movies and their views from movie viewers. The scores that IMDb gives are highly recognized by the audiences which represent the quality of content as well as audience's favor towards it. Therefore, in this project, we will try to unveil some of the important factors influencing the IMDb score and propose an efficient approach to predict success and failure of movie. In recent years, movie ratings are

influenced by several factors such as actor, story, release date, likes by viewers, etc. These factors produces proper prediction about ratings for the new movies being released. There also have been numerous semantic analysis techniques to analyse user reviews which were applied to analyse the IMDb movie score. In this project, we use the IMDb dataset to predict whether the movie has made a profound impact on our society or not. Cinema is one among the foremost powerful media for mass communication within the world. Cinema has the capability to influence society each regionally and globally. Many different kinds of movies are produced each year. Some movies portray historical events, some illustrates a culture, whereas some give fantasy, and some do many more. We have a tendency to perform associate exploratory analysis of the data and observe some attention grabbing phenomenon, which also helps us to make our prediction strategy better. Also we will come to know about the features which affect the IMDb score.

The first step is to identify a dataset of movie data which is suitable for analysis. Relevant attributes need to be selected from the movie data. Attributes can be general pre-production information regarding film productions such as movie title, sequel, genre, language and information about writers, actors, and directors. Similarly, the data must include some measure of success, such as user movie ratings. Secondly, the relevant dataset has to be prepared and structured in such a way that the data used is representative of the movie scene at large, as well as suitable for analysis by the relevant machine learning techniques and algorithms. Further, correlation is performed on relevant dataset to find the relationship between al the variable with each other. The important step in training oursystem is to apply classification model. There are many classifiers. Lastly, the prediction performance of the relevant machine learning algorithm has to be evaluated on the dataset in order to determine success and failure of movie accurately

Conclusion:

In this research, the systematic approach towards some efficient strategies for movie success and failure prediction is presented using machine learning. The IMDB is an interesting dataset to analyze. A movie success does not only depend on features related to movies. Number of audiences plays a very important role for a movie to become successful. Because the whole industry is about audiences. The whole industry will make no sense, if there are no audience to watch movie. Number of ticket sold during a specific year can indicate the number of audiences of that year.

By

P.Rupa