Indian Cinema Success Rate Analyzer using ML

Ansh Bhatia(008), Bhanu Soni(013) and Dhruv Soni(016)

Department of CSE, IET, JK Lakshmipat University

Abstract

The main purpose of this study is to develop a movie rating, movie box office, prediction model (Bollywood film) by being a Hit or Flop, long before the movie was released and a recommendation system. Using the **Regression algorithm** and **Classification algorithm**, the machine is monitored dependency algorithm, model designed for prediction Hit Bollywood movies. The advanced model was able to predict Hit movies with up to **60% accuracy**. We proposed model differs in the concept you use Rating of the movie as a prediction Bollywood movie feature.

Research has a lot to do with for staff and educators as possible in the beginning benefit from using this model to predict success a Bollywood film before its release and later may have a new find details in this field of knowledge.

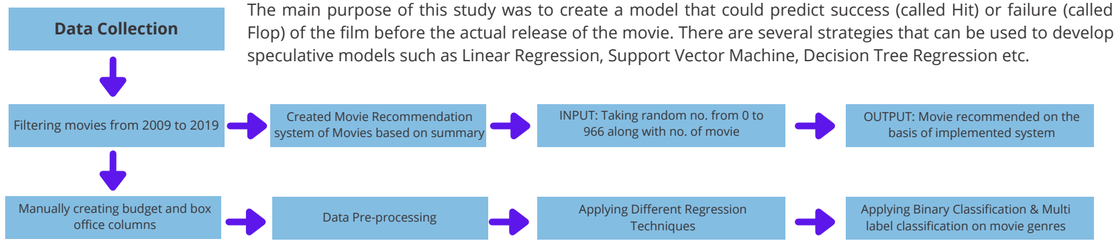
About the Dataset

We have chosen a **tabular data** that contains data of around **4329 movies** with **20 different parameters** like Movie Title, Year of Release, IMDB ratings, IMDB ID, runtime, genres etc.

Further, we processed(**data pre-processing**) the selected data according to our needs and **filtered** the movies from **2009 to 2019**. Hence, we removed the unimportant parameters to make the data more useful.

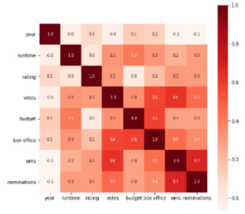
Reference link: **https://bit.ly/30jwj6n**

Approach

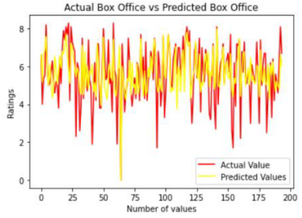


The main purpose of this study was to create a model that could predict success (called Hit) or failure (called Flop) of the film before the actual release of the movie. There are several strategies that can be used to develop speculative models such as Linear Regression, Support Vector Machine, Decision Tree Regression etc.

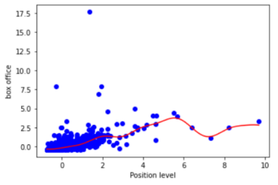
Co-Relation Graph



Actual vs Predicted Ratings



Support Vector Machine graph



Classification Table for Dataset

Algorithms Used		Target Variable = Box Office		Target Variable = Ratings	
		RMSE value	R sq. score	RMSE value	R sq. score
Linear Regression			0.38		
Random Forest Regressor					0.39
Decision Tree Regression		61.4		1.58	
Ridge Regression		0.44	0.19		
Lasso Regression		1.59	0.08	1.51	0.09
Multi Label Classification	Binary Relevance		0.66		
	Classifier Chain		0.16		
	Label Powerset		0.63		
kNN Algorithm		94.74		1.402	

Discussion

In this study, an effort was made to develop a simple system a model for predicting the success of a film before its release using the different **Regression** algorithms. Using **MovieVerdict** (Hit or Flop) as a result variable, the model used three predictions **Movie Ratings, Movie Genres, Movie Budget** a film as **predictive variables**. It was found that three Selected predictions have great **interpretive power** find the **success or failure** of a movie with Movie ratings has been found to have a major influence on the making of a hot film or flop, followed by the **IMDb rating**.

After all of this, we have provided the recommendation system for the movies i.e., if you want to see the related movies from one movie then you can use this recommendation system. We recommend a movie on the basis of the summary in which if most of the keywords are the same in the movie then It will recommend the movie which is similar.

References

1. Predicting success of Bollywood Movies using ML techniques by **S.R. Jaiswal** & **D. Sharma**
2. Predicting box-office success of motion pictures by **R. Sharda** & **D. Delen**
3. Pre-production forecasting of movie revenues with a dynamic artificial neural network by **M. Ghiassi**, **D. Lio** & **M. Brian**.
4. Research on the subject by **K Meenakshi**, **G Maragatham**, **Neha Agarwal** and **Ishitha Ghosh**