INTRO PAGE

I’m unable to remove page borders to all pages.

Either border or header-footer has to be used.

CERTIFICATE PAGE

ABSTRACT PAGE

ACKNOWLEDGEMENT

1

Introduction

* 1. Movie Prediction

Movie making involves huge investment thus movie prediction plays a vital role in the movie industry.

Movies is the most convenient way to entertain people. However only few movies get higher success and are ranked high. Many movies are produced by the movie industry in a year. A movie revenue depends on various components such as cast acting in a movie, director of the movie, film critics’ review, rating for the movie, genre of the movie, etc. Because of these multiple components there is no formula that helps us to provide analysis for predicting how much revenue a particular movie will be generating. However, by analysing the IMDB score generated by previous movies, a model can be built which can help us predict the expected quality for a particular movie. As we know in today’s world, the movie is one of the biggest source of entertainment and for business purposes. To expand this business further we need the technology through which we can predict the success rate of the movie. Success rate of movies, models and mechanisms can be used to predict the success of a movie. It will help the viewers whether to watch the movie or not as the quality of the movie will be predicted. Stakeholders such as actors, producers, director etc. can use these predictions to make more informed decisions. They can make the decision before the movie is released. This proposed work aims to develop a model based upon the data mining techniques that may help in predicting the success of a movie in advance thereby reducing certain level of uncertainty. The excellent way to find detailed information about almost every film ever made is through IMDB.

TO BE USED FOR SOME OTHER PURPOSE

Vast amount of data, which contains much valuable information about general trends in films. Data mining techniques enable us to uncover information which will both confirm or disprove common assumptions about movies, and also allow us to predict the success of a future film given select information about the film before its release. So here we are developing the software for data analytics through which we can predict the success rate of the movie with high accuracy. Here we are using the R-software to predicting the movie success rate into which first we have downloaded the data set from kaggle.com and after that we are generating the training and test data set. In a dataset, a training set is implemented to build up a model, while a test (or validation) set is to validate the model built. The main attributes selected for building model are critics\_score, imdb\_rating, imdb\_num\_votes, audience\_score. Data points in the training set are excluded from the test (validation) set. Usually, a dataset is divided into a training set, a validation set (some people use 'test set' instead) in each iteration, or divided into a training set, a validation set and a test set in each iteration. These training and test dataset is used to build model for selected set of attributes. On the basis of the generated model prediction have been done and result have been generated. Through the acquired result we can easily conclude that the movie is hit, superhit,

blockbuster or flop. The outcome of this research is therefore twofold, it provides tools and techniques to transform the database data into a format suitable for data mining, and provides a selection of information mined from this refined data.