



WAQAR AHMED
01-243181-025

Movies Quality Prediction Using Machine Learning Techniques

Masters of Science in Computer Science

Supervisor: Dr.Sabina Akhtar

Department of Computer Science
Bahria University, Islamabad

December 6, 2020



MS-13 Thesis Completion Certificate

Student Name: **Waqar Ahmed** Registration Number: **56113**

Program of Study: **Masters of Science in Computer Science**

Thesis Title: **Movies Quality Prediction Using Machine Learning Techniques**

It is to certify that the above student's thesis has been completed to my satisfaction and, to my belief, its standard is appropriate for submission for evaluation. I have also conducted plagiarism test of this thesis using HEC prescribed software and found similarity index at **12%** that is within the permissible set by the HEC. for MS/MPhil/PhD.

I have also found the thesis in a format recognized by the BU for MS/MPhil/PhD thesis.

Principle Supervisor's Signature: *Sabina*

Principle Supervisor's Name: **Dr.Sabina Akhtar**

October 7, 2020



MS-14A
Author's Declaration

I, **Waqar Ahmed** hereby state that my MS thesis titled "**Movies Quality Prediction Using Machine Learning Techniques**" is my own work and has not been submitted previously by me for taking any degree from "**Bahria University, Islamabad**" or anywhere else in the country / world.

At any time if my statement is found to be incorrect even after my Graduate the university has the right to withdraw cancel my MS degree.

WAQAR AHMED
01-243181-025
October 3, 2020



MS-14B Plagiarism Undertaking

I, Waqar Ahmed solemnly declare that research work presented in the thesis titled

Movies Quality Prediction Using Machine Learning Techniques

is solely my research work with no significant contribution from any other person. Small contribution / help whenever taken has been duly acknowledged and that complete thesis has been written by me.

I understand the zero tolerance policy of Bahria University and the Higher Education Commission of Pakistan towards plagiarism. Therefore, I as an author of the above titled thesis declare that no portion of my thesis has been plagiarised and any material used is properly referred / cited.

I undertake that if I am found guilty of any formal plagiarism in the above titled thesis even after award of MS degree, the university reserves the right to withdraw / revoke my MS degree and HEC and the university has the right to publish my name on HEC / University Website on which name of students who submitted plagiarised thesis are placed.

WAQAR AHMED
01-243181-025
October 3, 2020

Abstract

Quality prediction plays an important role in the business outcome of the product. Due to business interest of the concept, it has extensively been studied in the last few years. Advancement in computational intelligence techniques and with the advent of robust and sophisticated machine learning algorithms, it is require to analyze the factors influencing the success of the online products. Due to busy life routine people prefer to buy products online to save their time. one of the major issue that online buyer face is that how to identify the product quality online. There are many options that can be used to identify the quality of the product and their services. Reading peoples reviews on social media platforms and applying all of the concern decision attributes on algorithms to analyze the quality. Online business is all about the predictions. stock market and Forex traders predict the commodities and currency rates then they put their lots according to the prediction. similarly other online businesses use some predictions model to estimate their sales and purchases. They all use some predictions algorithms to compute their future events and the quality to arrange resources accordingly.

Movie industry is a huge market. Today, thousands of movies are released every year. This generates millions and billions dollar of revenue. There is a huge team and hard work behind each movie. The movie industry invest a heavy amount on their films to make it better and top in the box office. One of the film industry named as motion film industry situated in the United States that invest sixty millions dollar on each movie. With this huge amount of investment, they also predict good ROI. The best ROI is only possible when they make quality content which attract the viewers. When more viewers watch the movie, it will result to generate a good profit. The movie industry thoroughly study all of the concern factors that can influence the movie before starting its production. After production they give a little teaser to the viewers in terms of movie trailer and then decide when and how to release the movie. The main focus of each movie industry is on the content use in the movie and the actors that play their role. Since these two factors have a huge impact on any movie success. They also use their past experience to make their new movie more better to generate more revenue.

In this work, we will apply computational intelligence techniques such as Machine Learning algorithms to predict the quality of the movie using benchmark data set, i.e. IMDB, however, this work could easily be extended to other relevant fields like identifying the quality of products etc. Our data set consist of 1000 movies witch includes different distributed attributes. The movies quality prediction is done by using blogs and people reviews on social media platforms but very less work is done by using proper attributes and features used in the movies and apply some algorithms. The prediction is based on these decision features including movie title, actor, total numbers of downloads, views, likes, Movie budget, business amount, Rating, Votes Metascore, runtime etc. Here the Rating attribute is act as a label or class. Finally, we will label the prediction results into four classes from Flop to blockbuster. The movie that gives us very high rating will be considered as blockbuster movie and the movie that gives us very bad rating will be consider as flop movie.

The implementation is done in Anaconda in which python language is used and apply machine

learning algorithms. We have used four machine learning algorithms named as Linear Regression, Decision Tree, Random Forest and Naive Bayes to perform classification and we also identify which classifier gives the higher accuracy rate among the others. The results are categorised into four different classes. Flop, Average, Good, blockbuster. We used these algorithms and compute the results that shows, which algorithms gives us the best and higher accuracy score.

Acknowledgments

First and foremost, I would like to thank **ALLAH Almighty** for giving me the strength, knowledge, ability and opportunity to undertake this research study and to persevere and complete it satisfactorily. Without His blessings, this achievement would not have been possible.

I would first like to thank my supervisor, **Dr. Sabina Akhtar**, who guided me in my journey towards this degree. The door to Dr. Sabina Akhtar office was always open whenever I ran into a trouble spot or had a question about my research or writing. She has been there providing her heartfelt support and guidance at all times and has given me invaluable guidance, inspiration and suggestions in my quest for knowledge. Without her able guidance, this thesis would not have been possible and I shall eternally be grateful to her for her assistance.

My acknowledgement would be incomplete without thanking the biggest source of my strength, my family. The blessings of my parents and the love and care of my siblings. I would also like to thank my friend **Mr. Syed Jawad Ali** for his care and love throughout this journey. This would not have been possible without my family and Friend unwavering and unselfish love and support given to me at all times.

WAQAR AHMED
Islamabad, Pakistan
December 6, 2020

*“All we have to decide is what to do
with the time that is given us.”*

J.R.R Tolkien,
The Fellowship of the Ring, 1954

Contents

1	Introduction	1
1.1	Research Background/Overview	2
1.2	Research Objectives	4
1.3	Motivation	5
1.4	Research Contribution	6
2	Literature Review	9
2.1	Naive Bayes, Bagging, J48, Random Forest and SVM	9
2.2	Collaborative Filtering, Decision tree, Artificial bee colony	11
2.3	Regression algorithms, Visualization Paradigms, SVM	14
2.4	Neural Network Used in the following Research's	16
2.5	Other Relevant Algorithms that Used to Compute Movies Quality	17
3	Methodology	21
3.1	Implementation steps	22
3.1.1	Selection of Movies Attributes	22
3.1.2	Scaling of extracted features using Training data set	22
3.1.3	Applying classification algorithm	23
3.1.4	Developing threshold value	23
3.1.5	Testing classification results with testing data set	23
3.1.6	Evaluating calculated results with developed threshold value	23
3.1.7	Terminate the algorithm	23
4	Implementation and Results	25
4.1	System Architecture	25
4.1.1	Data Extraction	26
4.1.2	Feature Selection	26
4.1.3	Data Preprocessing	26
4.1.4	Apply machine learning algorithms	27
4.2	Machine Learning Classification Algorithm	28
4.2.1	Linear Regression	28
4.2.2	Decision Tree	28
4.2.3	Random Forest	28
4.2.4	Naive Bayes	28
4.3	Libraries	29
4.4	Data set	29
4.5	Results of our proposed model	29

5	Conclusions	37
5.1	Future Works	38
	References	39

List of Figures

3.1	Implementation Steps	24
4.1	System Architecture	27
4.2	Effect of Metascore on Rating	31
4.3	Effect of Votes on Rating	32
4.4	Effect of Revenue Generated by Movie on Rating	33
4.5	Effect of Runtime on Rating	34
4.6	Effect of Combination of Metascore & Votes on Rating	35

List of Tables

2.1	Summary of Naive Bayes, Bagging, J48, Random Forest and SVM	11
2.2	Summary of Collaborative Filtering, Decision tree, Artificial bee colony	13
2.3	Summary of Regression algorithms, Visualization Paradigms, SVM	16
2.4	Summary of Neural Network Used in the Research	17
2.5	Summary of Other Relevant Algorithms	19
3.1	Class of Movies Consideration	22
4.1	List of Classifiers Used in Proposed Model	28

Acronyms and Abbreviations

IMDB	Internet Movie Database
SVM	Support Vector Machine
ML	Machine Learning
HD	High Definition
ARSA	Auto Regressive Sentiment Aware
ANFIS	Adaptive Neuro Fuzzy Inference System
DBSCAN	Density Based Special Clustering of application with noise
NB	Naive Bayes
ABC	Artificial Bee Colony
SMOTE	Synthetic Minority Over-Sampling Technique
SGD	Stochastic Gradient Descent
ROI	Return On Investment

Chapter 1

Introduction

Predicting quality of the product is a challenging task in data mining domain. The prediction is based on the past and present features of the product. Machine learning algorithms are used to Predict the Movies quality in terms of Flop, Average, Good and Blockbuster. In our proposed work We have used Hollywood movies data set. Machine learning algorithms were trained using training data set and then tested using testing data set to assure the accuracy of the predicted results. The machine learning algorithms classify the movies quality from blockbuster to flop. Predicting quality of the movie is the fundamental problem in machine learning domain [1]. There are different machine learning algorithms that can be used to predict the movie success. The machine learning algorithms use computational intelligence techniques to forecast the future events. It has the ability to learn from the past experience and from historic data set.

Predicting quality of the product have become more common today. People always choose the products by using some quality predictions model. There are many prediction model in online business that suggest the quality of the product to the end user. Some people use their own mind at real time and predict the quality of the product. There are some people who properly use tool and techniques to predict quality of the product. Prediction model most of the time give accurate or nearly accurate results as it use all the concern distributed features that required in prediction. The prediction model is well trained to forecast the future events because it use the past experience and traditional data for predicting the quality of the product. The model classifies the product in different categories i.e. high quality, average quality and low quality. The user chooses the product on the basis of result given by the quality prediction model.

In movies quality prediction the prediction models help both the user and the movie maker. It helps the user to choose a specific Genre of movie to watch i.e. thriller, Adventure, Action, romance, war etc with highly rated movie of the week, month or year. It also help the movie maker that what kind of content and resources will generate more revenue on the box office business. There are three different kinds of prediction approaches [2]. Content based approach, collaborating filtering and hybrid methods. Content based approach predict the quality of the movie on the basis

of user reviews about the content used in the movie. User rates the content use in the movie whether they like the content or not. Collaborative filtering approach predict the quality on the basis of similar users' reviews. It classifies the movies in different categories with same genre of movie and allocate the category to the similar user of interest. Some people like action and adventure while some like romantic and comedy. The hybrid method combines both of the above-mentioned approaches. Accurate rating can only be calculated by using correct and effective data set. The introduction describe in following headings.

- Research Background/Overview
- Research Objectives
- Motivation
- Research Contribution

1.1 Research Background/Overview

The quality predictions solve many problems as it recommends the end user that what to choose and what is best. People get confuse by selecting a product from given bunch of data that cause information overload. This kind of problem can easily be solved by the prediction systems. The prediction model also help the user to save the time in selecting a product. The prediction system recommends a specific movie to a specific user of his own choice. It assess the user that what kind of movie he would like to watch by analyzing his previous reviews and ratings. The prediction model also helps the movie maker in targeting a specific type of user and specific type of script in the movie. Targeting a specific movie for a specific user leads to generate more revenue on the box office business.

The movies quality prediction helps the movie maker and creator to improve the quality of content in the movies. For example, if a movie generate less revenue at box office and get negative review from the user then the movie makers highlight those week points in the movie and work on these points to overcome the problems. They make content of the movie in which people took more interest, as a result it will generate more revenue on box office. So, if the movie generates more revenue and also get positive review from the user then it will be predict as a high-quality movie or a Blockbuster movie. The external factors also influence the movies revenue for example, movie release date and weather etc. Some movies release on some special and traditional day of the country. On these special occasions people usually go outside to have some fun after getting bored in daily routine life schedules. Some of them specially goes to cinema houses and watch movies for entertaining. Most of the movie's release on these special occasions which leads to generate more revenue than normal release days. So, these external factors matter a lot in prediction system to get the accurate result.

Since we use the machine learning classification algorithms to obtain movies class whether the movie is flop, average, good or Blockbuster. Collaborative filtering [3] plays an important role in

classification. There are hundred of movies available on the internet so the people get confuse to select a movie to watch which also consume a lot time of the people by selecting a single movie from a whole bunch of Movies data set. Collaborative filtering techniques are used to filter out high rated movie to the user. Basically, it saves time of the user in selecting a specific movie from a whole bunch of available movies. Collaborative filtering does not consider the content of the movie it only considers the interest of the similar user in any genre of the movie. Different People like different movie genre for example some people like war, thriller and action movies and some like romantic, comedy types movies. The collaborative filtering approach is use to filter the same types of movies to the same types of people. Basically, it classifies different genre of the movies and allocate the same genre to the similar interest of the user.

Online reviews directly influence the revenue of the product. It matters a lot to both the end user and the seller. We cannot measure the quality of the movie only by the user reviews because sometime movie generates high revenue but gets negative review from the users. The social media platforms play an important role to create hype. The movie may not be a great but only because of that hype created by the social media platform, people gets crazy to watch the movie and the movie generate high revenue. We have used hybrid features to measure the quality of the movie. We set a threshold value in the form of movies rating feature which used as label. The labels are classified in four different categories whether flop, average, good or Blockbuster. So movies will belong to the particular class after computational results. The user like the content use in the high-quality movies so it gets positives reviews. The user only took interest in a high-quality movie and it also generate more revenue in the box office business.

Sometime a very new user wants to watch a movie. The system does not know his interest because he is a new user and he has not participated in any kind of reviews and feedback. Similarly, when a new movie come into box office market the movie also not get any reviews from the user yet. The solution for the new user is to recommend the user some popular and interesting movies then gets reviews and feedback from that user about the movie to know his interest. After that the system will know the user interest whether he likes fiction, thrilling, war or romantic movies etc. The similar user when want to see next time any movie the system will generate list of recommended movies to that user. The best example is YouTube. When we search and watch anything on YouTube, the YouTube generate Hundreds of similar videos for us. The movies success depends on the profit that movie generate rather than the investment on the movie. Sometimes high budgeted movies get flop and similarly sometime very low budget movies create hype and get houseful shows and hit on the box office business. So, the movie success is purely depending on the Return on investment that is generated by the movies.

Online market is growing very fast. There are some popular examples like Amazon, Alibaba and Daraz have a huge online business. People prefer to buy product online. They have no time to go to the shops because of busy life routine. One of the major drawbacks of the online shopping is the product quality. There are many complains of the people that the stuff they receive at home was not the same they had selected to buy at online portal. So, the people use reviews given by the end user on different social media platform to analyze the product quality and their services.

Social media platform plays a vital role in either product gets flop or generates high income. For example, when a movie gets viral People took interest in the movie on the basis of those review given by the user on Twitter, Instagram, Facebook and other social media platform. No one knows the actual crew member of the movies like actor, director, etc. just because of that hype created by the social media, so people get curious in that viral movie and took interest to watch it. These are some external factors that also impact on the box office business.

The information is increasing rapidly over the internet than the ability of human processing speed. The Recommender system plays a vital role to process bulk of information into a well specified design. There is bulk of information on the internet which also include some raw data as well as important materials. This mixture of data consumes a lot of time of the users to filter out their desire data from a bulk of given information by using their own processing ability. Considering the movie recommender system, it's hard to extract a movie manually from a given data set consisting of hundreds and thousands of movies. To extract our desire movie or a category of movie i.e. different genre of movie like action, comedy, thriller etc. we use some algorithmic approach. The more efficient algorithm will be the most accurate and reliable results we will get.

1.2 Research Objectives

Quality prediction plays an important role in the business outcome of the product. Due to business interest of the concept, it has extensively been studied in the last few years. It uses some recommendation system techniques that can automatically generate high quality product to the end user based on user preferences. The data consist of raw materials as well as useful information in it. Recommender systems are those system that can easily filter out important data to the users from a given bunch of information overload. Movie quality prediction has a lot of use in the business field and based on this prediction movie production companies can plan their resources accordingly. For example, if movie quality prediction come to be true, it will lead to increasing in revenue for company. Quality prediction is very helpful for the box office business and also for those people who are investing in share market. The main objective is to saves the time of the end user by applying machine learning algorithms by ranking the product quality to the end user. It can also be helpful to product owners that they can improve their products and services according to the user needs.

Movie prediction systems have a lot of use for different companies to plan their resources in an efficient way. For example, The Movie studio expects that, its new coming movie will be a high quality movie by analyzing their past experience, so they will rent more movie theatre in advance. The more people will be able to watch the movie and that will generate more revenue on box office. If they rent less theatre room in advance, not all viewers might have been able to watch the movie in its opening weekend, it will directly influence movie revenue Box Office. In movie industry the opening week for any new upcoming movie is matter a lot. When that first week after the movie release makes a houseful show at cinema houses, then it will generate a good profit to the film industry. Some of the most famous movie name as the Avengers and star war generate very good

profit in the opening week. Avengers End game season generates \$357,115,007 in just the first three days of their opening week which also cover the cost of movie with profit. Similarly, Star Wars the Force Awakens generates \$247,966,675 dollar in just three days of the opening week, which also cover the cost of movie with profit.

Today, many prediction systems are available for predicting movie quality based on decision features such as viewer ratings and box office incomes [4] in the first week of movie release. For predicting the quality of the movie, the main criteria is the quality that how a user rated the particular movie and its revenue and some other external factors such as release date, competing movie and weather also influence the movie quality because it affects revenue of the movie. The past data are very useful for predicting movie quality even before its release. For example, if a movie of some actors and directors give a houseful show in the first week of their movie release then there is a chance that their next upcoming movie can also be a super hit. There are some movies that are released in parts or with the same cast, but with different seasons, for example Avengers comes up Five times with the same cast and with the same story with leading to the previous story. This movie is the top movie in generating revenue in the first week of the movie release. The movie studios make advance booking of cinema houses to make it possible that many people can get the chance to watch the movie in its opening week. So, in this way they can make a good return on their investment in the first week.

1.3 Motivation

Due to the rapid increase of information over the internet it is difficult for the user to select a high-quality product. Data is generating very fast over the internet and thousands of movies are produced every year. This rapid increase of data change the whole scenario of the online business. A lot of people have their own channels on YouTube and other social media platforms. They produce their self-made content without any proper knowledge about the story and upload it on their channels. The quality of the content matters a lot. The popularity of the channel and the revenue generation is depending on the content quality and the number of viewers. Good quality content attracts more viewers to the channel which leads the channel in high ranking. There is no way to copy other contents because of the strict policies apply by the social medium administrations. So each channel tries to create their own unique content. People take an interest in giving feedback and reviews. These reviews play an important role in the success of business. The new user read the review of the old user and then decide whether to use the product or not. These reviews also play an important role for the owner of the product. It identifies the user priorities and highlights the flaws in the product which make easier for the owner to resolve these highlighted issues identify by the end user.

Quality prediction sets new trends in the world of online business. When multiple peoples give review, it matters a lot instead of getting reviews from a single person. More reviews from the end user gives more accurate results about the product quality. These reviews help the new user to take a final decision about the product by evaluating those reviews given by the older users. In

this way it will also help the owner to improve product quality with time by analyzing the end user feedback. There are many quality prediction algorithms which are used by the peoples in evaluating the product or service quality. The reviews given by the old user are used to train the algorithms which make the algorithms more effective and efficient. The best example is when we visit to the branded hotels after serving the food, they distribute the feedback booklet to each customer to get the reviews about their food and services from each individual person. In movies quality prediction when a movie gets positive review from the viewers then more people show interest to watch the movie. The new user will prefer to watch that specific movie because of reading those positive reviews given by the other viewers. Content used in movies is evaluated by the number of views, number of downloads, number of likes and dislikes, reviews given by viewers, Revenue generated by the movie and actors and directors' efforts etc.

Movie quality prediction is closely related to the movie recommendation system to the users. Researcher proposed different models of machine learning algorithms for the user which uses some movie related attributes and ranked movies to the users. Many researches have already been done in this field. It's hard to find out a single movie from a given data set consisting of thousands of movies. The algorithm ranks all these movies into different categories starting from Blockbuster movie to flop movie. Obviously, the algorithm will recommend the user to go to Blockbuster movie. In this way these algorithms saves a lot of time of the user and effectively compute the movie quality. These algorithms can also be used for categories the movies in different genre. The algorithms are implemented on movies trailer or movie poster to categories the movie into a specific genre. There are many genres in movies studio departments, for example Action, Horror, Comedy, Fiction, Romance, Drama, crime film and animated movies etc. Each user has different taste in movies, some people like action, horror, crime films and some like romance and comedy type movies. This can also be useful to save time of the viewers by classifying the movie in different genre. The user will simply go to the specific genre and will choose their desire movie to watch. In this paper, we will use hybrid features to predict the quality of the movies based on downloads, views, user reviews and revenue generated on box office.

1.4 Research Contribution

The predicting quality of the movie is the fundamental problem in data mining domain. There are thousands of movies releasing every year. Selecting a high quality movie from a given bunch of movies data set is an interesting topic in the data mining domain. As we know that quality prediction plays an important role in the business outcome of the product. For predicting the quality of the movie, the main criteria is the quality that how the users rated the particular movie and its revenue and some external factors such as release date, competing movie and weather etc. All these external factors influence revenue of the movie. It is required to analyze the factors influencing the success of the movie. Good prediction system requires the correct interpretation of the available data about the movies. Therefore, for this problem, we use modern machine learning techniques to predict the quality of the movie, which will help movie production companies in order to plan

their resources accordingly such as revenue and release date planning etc. This will also help the viewers in selecting a high quality film efficiently and effectively. This work will make efficient and accurate predictions for people and business about a particular movie quality that whether it is a flop or a Blockbuster movie.

Our proposed model focus on the machine learning Algorithms [5] to predict the movies quality. We have used four machine learning algorithms named as Linear Regression, Decision Tree, Random forest and Naive Bayes. We have applied machine learning algorithm to predict the quality of the movie using data set: Movie IMDB [6] database. The data set consist of 1000 movies that contain distributed attributes used in the movie. We obtain a data set from a very well known data set repository called as kaggle. We have split the data set and use 800 movies in the training phase and the 200 movies to test the model. This technique will enable the prediction to make accurate and efficient recommendations to users. In our work, we will make efficient movie quality prediction based on hybrid features, which will be extracted from a data set. After extraction of features, we will apply Machine Learning Algorithms to label the predicted results from flop to Blockbuster movie. The results are classified in four different classes such as Flop, Average, Good and Blockbuster movie. We have used four machine learning algorithm and we have computed that which algorithm gives us the higher accuracy score.

Chapter 2

Literature Review

In this section, we will discuss some background information and related work for movie prediction systems. Neural network [7] is widely used in weather and other prediction systems. It forecast the future events. It works same like the human nervous system. Basically its a computational model that use some mathematical data and perform operation on it accordingly. A large interest has started in predicting movie success when Netflix announced the Netflix Prize [8] in October 2006. They use machine learning algorithms and data mining approach to predict the movie quality. They offer an open competition that whoever improve the accuracy of movie quality prediction up to 10% will win the Netflix prize. Their existing system of quality prediction was Cinematch. They offer prize of one million dolar to the winner who will make improvement in their existing prediction system up to 10%. In this challenge, different groups proposed different techniques to predict movie success. Netflix is a big online streaming platform, where users pay a monthly fee to watch movies and series on demand. They have full HD and best quality videos for their users.

2.1 Naive Bayes, Bagging, J48, Random Forest and SVM

Warda Bristi, [9] Film industry is not Just an entertainment industry. It becomes a huge business. The film industry make huge investment with collaborating other investors. Each movie industry work a lot on their new upcoming movies. There is a big team and huge efforts behind any film to make it better and successful. The author use machine learning classification algorithm to efficiently predict the movie ratted. They use input feature as studio, actor, director, screening, and genre. They create their data set by its own. They get Hollywood movies list from Wikipedia and rating of movies from IMDB. IMDB is the online movies platform where people gives reviews about the movies. IMDB use rating starting from 1-10. The movies that are released in 2018 are selected. 250 movies are collected to perform the experiment. On basis of IMDB rating the movie will be classified. The movie will be classify in four categories depending on IMDB movie rating. The movie could either be Flop, Below Average, Average or Hit. They use weka to appnly machine

learning algorithm. There are five machine learning classification algorithms are used named as Bagging, Random Forest, J48, K-Nearest Neighbour and Naive Bayes. Random Forest and J48 gives the best accuracy score as compare to other machine learning classification algorithms. SMOTE Technique is used to assemble that data because data was imbalance.

Nahid Qader [10] uses seven machine learning techniques. Out of these seven, multilayer perceptron neural network gives the best score in the prediction. They use IMDB dataset which consist of 755 movies. Initially the dataset was consist of more than 3000 movies but after filtering the dataset they left with 755 movies. The prediction model was not a binary. The movie rated from flop to excellent or blockbuster. They use five rating categories namely flop, average, good, excellent and blockbuster. The accuracy rate of MLP was 58.5% which was the highest score among other machine learning techniques. The reason of getting higher accuracy rate in Multiplier perceptron neural network is the dataset was complex and cannot be easily categories. The second higher accuracy rate was achieve by the support vector machine classification algorithm. The performance is evaluated by using numbers of audience or the tickets that are sold for a particular movie. They did not consider genre or sequel of the movie. The experimental results are evaluates by using 10 fold validation. The results are also validate by precision and recall function which shows that the model predict the results accurately. The movies success is predicted by using both of the pre-release features and the post-release features. The post release features only include the first week after the movies release date.

Rui Yao [11] Due to increase in online business, now the online reviews getting from customers is very common. Its changes the whole scenario of E-Commerce Business. It helps both the seller and the customers. In this paper they use reviews given by the user on social media platform like YouTube, Facebook, Twitter and Instagram etc. to predict the revenue of the movie. Using sentiment analysis approach and applying machine learning algorithm on the review to predict the movie revenue. They used 14 keywords that are used by the reviewers at the time of giving feedback on movies performance. for example these Keywords include flop, good, poor, super hit etc. They use only top three movies on which reviews given by the user. The movies that are used in research are The Amazing Spiderman, Life of Pi and Madagascar 3. There are 911 record in their data set. They used two types of data in their models. First they use previous sales data of the movies and apply it on their proposed algorithms which gives not a good results. Then they use sentiment data of the users reviews obtained from social media along with previous sales data. They apply both of this data on their proposed model which gives very best results. The Naive Bayes classifier is used to determine the expected box office revenue of the movies. Naive Bayes gives accuracy score of 78.75%.

P. Nagamma [12] They use sentiment analysis approach on machine learning classification algorithm called SVM and NB. Use TF-IDF approach to perform sentiment analysis. They select the feature from the data set and apply machine learning classification approach on both negative and positive review given by the user. The model work as follow, pre-process the data, remove stopword and perform tokenization. After selecting features use classification algorithm and evaluate the result. They take a single movie 'anabelle' reviews and perform sentiment analysis. They use

total of 165 reviews given by the users on that movie. They collect only 30 days reviews. They also include movie sales as one of the attribute along with people reviews. The experiments and implementations are perform in MATLAB. A machine learning algorithm called Support Vector Machine is used to predict movie revenue on box office business. sentiment analysis is used to know the people emotions and their opinion about the product. Its very helpful for the seller to update their stuff according to the people opinion. They apply clustering technique after pre-processing the data set. K-means and DBSACN Clustering techniques are used. Then they apply machine learning algorithms to classify the results. SVM gives 62% accuracy with clustering. Naive Bayes gives accuracy score of 72.5% with clustering. The accuracy score is computed with 10-fold cross validation.

Table 2.1: Summary of Naive Bayes, Bagging, J48, Random Forest and SVM

Ref	Technique Use	Tool Use	Data Set	Accuracy
[9]	Naive Bayes, Bagging J48, Random Forest.	Weka 3.8.3 tool	IMDB Dataset	80%
[10]	Naive Bayes Logistic Regression Support Vector Machine Random Forest AdaBoost, SGD Multilayer Perceptron Neural Network	python library Scikit Learn	IMDB Dataset	58.53%
[11]	Naive Bayes classifier, Support Vector Machine	MatLab	IMDB & Mojo	78.75%
[12]	Naive Bayes classifier, Support Vector Machine. K-means DBSACN Clustering	MatLab	IMDB	SVM: 62% NB 72.5%

2.2 Collaborative Filtering, Decision tree, Artificial bee colony

Mladen Marovic [13] It get easier for the user to choose an item from a given bunch of data using any of the recommender system. Machine learning algorithms and data mining approaches are used to mine the data and extract a useful information from it. In this research they Uses IMDB movie data set to determine a possible user rating for a specific item based on their proposed features. The data set consist of 9428 movies which include 6 different parameters. This paper uses both the user and particular item features in prediction process. The three types f approaches are used: content based approach, collaborative filtering and hybrid features approach. It uses content based methods with the following features taken from data set: genres, actors, directors and screenwriters.

Collaborative filter is used to filter out the users with similar type of interest in the movies. For example there are many genre of movies in which some like fiction, thriller, war and some users like comedy, romance and story types movies. The movie features used in the training process are genres, actors and directors. A user that give his reviews or rating about particular items can also be the same review or rating to the same type of other items. This type of scenario is called automatic rating generation about the same type of product for similar interest of users. K-nearest neighbour algorithm is used to find the similar interest of different user in same types of movies.

P. Li and S. Yamada. [14] There are many categories of movie which name as genre for example action, horror, comedy, science fiction, romance, animation, crime film etc. Each user have different taste in movies. In this research the author make a decision tree to predict that which user prefer which genre of movie i.e. Horror, Romance, Thrilling or War etc. Recommender system plays an important role to give the users a better option to choose. It uses collaborative filtering approach to know interest of the same type of people. There are different people who have similar type of interest. On based of decision Tree it categorises the similar interest of people in one class. This approach is called LikeMind approach which is used by many develop business organization. They highlights and then give solution to the sparsity, scalability and transparency problem. The two types of attributes are used, content type attributes i.e. its shows movie genre war, comedy, romantic etc. The other attributes are taken credit based attributes, it shows movies sales and budget and revenue. The work consist of three steps. In first step they make a decision tree. In the next step they analyze a user preference on the basis of decision tree by using well known inductive learning algorithms. Finally they get the result of data set by using classification. Total of 15 items are used in evaluation. Out of these, 10 was used in training phase and 5 used in testing phase. Movie lens data set is used in this research.

Hengsong Tan [15] collaborative filtering approach is used to predict the interest of customers in specific product. Collaborative filtering is not just used in textual data but also it is used in non-textual information. Today data is generating very fast, which result into the sparsity. Customers review play an important role in predicting the quality of the product. So for this purpose collaborative filtering is used. It removes the sparsity problem and give better result. They use movie lens data set in their research to perform their experiment and implement the proposed model. The data set is obtain from the group lens research project. The data set consist of 1682 movies. 943 user give review on this data set. These user give approximately one lac rating on these movies. Each user gives average of 20 rating on each movie. Statistical accuracy measure approach is used to evaluate the performance of the method. The proposed collaborative filtering approach gives best results. In the previous work the results obtain from the traditional collaborative filtering method was not as much better as compare to this research. The results are compared by with traditional collaborative filtering approach by using mean square error method.

Rahul Katraya [16] Recommender systems are used to allocate particular types of products to some specified users. This system suggest the user about the quality of the product by which user can take decision accordingly. The past reviews of the user are used in collaborative filtering to recommend particular type of the movies to similar interest of users. If a user took interest in

Comedy movies so the next time the system will recommend new comedy movies with similar content to the same user. The author proposed hybrid recommender system. In this model he use K-means Clustering along with artificial bee colony. The proposed model is applied on movielens data set. The model gives best accurate results as compare to existing approaches. The accuracy of the proposed model is determine by applying precision and recall. The proposed model is implemented on a simple configured system easily available to any user. The system configuration is i3 processor, 1.9 ghz and 8GB RAM.

Zan Wang [17] There is bunch of data available over the internet. This large amount of data cause information overflow. This data contain both raw data as well as some useful information. It is very time consuming to extract important information from a given bunch of data because this data also include irrelevant information. For that kind of problems a recommender system is used. This system allocate data to the end user simply by knowing his/her interest. In movies quality prediction, a data set contain thousand of movies. Recommender system filter out movies to the user according to his/her preferences. It is very helpful to the user. Instead of searching one by one to select a desire movie the recommender system ranks all movies and give it to the end user. In this research, the author proposed Combination of k-means clustering algorithms along with Genetic algorithm. K-means is the most simplest clustering algorithm that iteratively work and choose the most nearest neighbour to it. It perform clustering by taking arithmetic means between two of its neighbour. At start we choose a centroid and then perform further computation. Genetic Algorithm is the adaptive version of natural theory. It uses chromosome model. Chromosome compute similarity of items to distinguish the fit and unfitt items. On the basis of fitness it create new items similar to the old one. They use Principle Component analysis technique in his work to perform data reduction. Movie lens data set is used to perform the experiment.

Table 2.2: Summary of Collaborative Filtering, Decision tree, Artificial bee colony

Ref	Techniques Use	Tool Use	Dataset	Accuracy
[13]	Collaborative methods Content based methods Hybrid methods	Machine Learning Data Mining Tool and Techniques are used.	IMDB	—
[14]	Collaborative filtering Decision tree Contents-based filtering Inductive Learning Algorithm	—	Movielens	63%
[15]	Collaborative filtering	—	Movielens	—
[16]	Collaborative filtering K-means Clustering Artificial bee colony	Core-i3 processor, 1.9 ghz and 8GB RAM Data filtering tools	Movielens	53.22%

[17]	Collaborative filtering, k-means clustering Genetic algorithm, PCA Computational Intelligence.	Dual Xeon 3.0GHz, 8.0GB RAM Matlab R2011b	Movielens	78%
------	--	---	-----------	-----

2.3 Regression algorithms, Visualization Paradigms, SVM

Mennatallah El-Assady [18] They proposed a hybrid features based model to predict movie success. Hybrid features are the mixture of features that impact directly or indirectly on the movie success. They divided their proposed features into two groups (i) movie internal features and (ii) movie external features. The first group consists of all internal features of the movie such as title, actor, actress, director, writer and genre etc. All these features are used to determine the movie itself and can influence the movie itself because movie quality depends on all these intrinsic features. The second group consists of all external features of the movie such as weather conditions, release date, competing movie etc. All these features do not influence movie quality but can influence movie revenue at Box Office. Some movies release on some special occasion when people goes to the cinema houses and watch the movies. This factor impact on movie revenue without considering the movie quality. They use IMDB movie data set and apply machine learning algorithms to predict movie quality. The data set contains movie budget, rating, and the runtime of the movie. They also perform sentiment analysis by obtaining user views in the form of comments or reviews given on social media platforms.

Robert Krueger [19] Designed an iterative approach to predict movie success. Using iterative approach, data is collected iteratively from the data set and implements machine learning algorithms to perform task. The analyst can select features based on their importance, which he/she thinks will be most important for the movie to predict and also a prediction model. The following machine learning algorithms are used to obtain their results Support Vector Machine, Multi Layer Perceptron or a logistic regression model. Furthermore, he/she gets feedback via error measures by predicting historical data with the currently selected model. They used Twitter data set for exploring features via sentiment analysis. They use IMDB data set along with sentiment data analysis obtain from twitter. The data set contain only those movies which are releases in 2013 in United States. The revenue generated from past movie in first week of the release in which same actors play their role is used to predict the movie rating of upcoming new movie. In this way they compute movie revenue of the new movie. Twitter comments of the viewers are used that are generated by the viewers just before fourteen days of the movie release. They perform sentiment analysis on these tweets and comments to obtain movie rating.

Dr.R.R.Sedamka [20] The marketing strategy is totally changed all over the world. The new trends of online marketing or advertisement is discovered. This digital marketing strategy saves a lot of time and resources of the users as well as the owners. In movies industry a movie can be a

flop or all time block buster. There is huge investment on each movie and when any movie gets flop it make loss of million of dolar's to movie industry. Similarly if a movie becomes super hit or block buster then it gain million of dolar profit to the movie industry. They proposed auto regression methods and adaptive network of Fuzzy inference system to predict movies Quality. In this work, they have implemented ARSA model to predict performance. It takes two inputs (i) The first one is an online sentiment rating extracted from blogs using sentiment analyzer and (ii) The second input is box office revenue. They implemented their proposed model on these two inputs and get single output which is category of the movie such as Disaster to All Time Blockbuster. They categorize the results into Eight different categories, Disaster, Flop, below Average, Average, Hit, Super hit, Block Buster and All time Block buster. This categorization is based on movies rating and revenue. The data set was consisting of only 145 movies. These movies were released in between 2010 to 2013. The accuracy of the model is measured by the mean square error. ANFIS model gives better accuracy for predicting movies rating and revenue as compare to ARSA model.

Muhammad Hassan Latif [21] uses IMDB Dataset to predict the popularity of movie using machine learning techniques. His main focus for predicting the movie quality was on the main crew member of the movie like actors, directors genre rather than social media networks like YouTube, Facebook reviews as sentiment analysis. There can be more than one genre in a single movie for example some movie shows only romance so its means it have only one genre whereas some movies shows romance, war, comedy etc. so it is consider as multiple genre movies. The movie popularity ranked from 0-10. 10 have the highest popular movie and 0 refer as terrible movie. If the rating is greater than 5 and less than 7.5 then it will be considered as average movies and vice versa. The proposed method consist of six machine learning classifications algorithms. Logistic Regression, Simple Logistic, Multilayer Perceptron, Navie Bayses, J48 and PART. The accuracy rate of simple logistic regression was higher than all other classifiers. Simple logistic give 81.34% and logistic regression give accuracy score of 81.15%. The main features that are consider to perform the calculation was user ratted, award that receive the movie and numbers of shows.

Subramaniaswamy V [22] use multilinear regression and SVM to predict movie success. They use movie trailer, release date and rating given by the user as a parameters to predict the movie success. They use dataset Boxoffice mojo and classify dataset into three categories low budget, medium budget and high budget movie. The reviews were collected from the YouTube. The dataset file is transferred to simple Excel sheet. Fourteen features were collected from the dataset as attributes. Each feature is arranged in a separate Colum of the excel sheet. The accuracy rate of classification using SVM was 56.25%. Support Vector Machine use as a classification algorithms. It categorizes the similar data in one class. Basically it make classes of the similar data. They take people reviews from the social media instead of movies internal features. Sometime words of mouth play an important in a role product success. Sometime a lower budget movie gives more revenue than the higher budget movie. Its means that review of the people in on the social media platform play more important role instead of considering the movie budget. The implementation of the model is done on Weka.

Michael Lash [23] their main focus was on the cast, content and release time of the movies.

They predict the movie success prior to the production of the movie. The proposed model will compute that whether the movie will earn profit or not on the basis of reviews on movie trailers by the user and the movie content. The proposed model will compute the movie success or failure before the production of the movie. This model will help the investors and the movie makers to make some changes before finalizing the production of movie. So the movie will earn more profit. The Box office mojo show the number of release days of the movie and how much the movie earn in these days. It also shows the information about the total investment on the movie and the profit that earn by the movie. In this paper the information is gathered from box office mojo website. The dataset contains 1353 movies. In this dataset the movie budget and the movie revenue is included. The prediction prove that out of 1353 movies 384 movies will be profitable. The model separately compute the accuracy of each class or category. The parameters were categories in three classes. The model use only the content of the movie and perform prediction. The model use only the cast of movies and perform prediction. The model use only the release date and time of the movie and perform prediction. The sperate results were obtained for each class. The main aim of the work was to predict the movie success in terms of profit gain in very early stages of the movie production.

Table 2.3: Summary of Regression algorithms, Visualization Paradigms, SVM

Ref.	Techniques Use	Tool Use	Dataset	Accuracy
[18]	Multinomial Regressions visualization paradigms	Python script, MySQL	IMDB	69%
[19]	Linear Regression Support Vector Machine Multilayer Perceptron	Feature lens	IMDB	75%
[20]	Auto regression Fuzzy inference system	Python	IMDB	78%
[21]	Logistic Regression Multilayer Perceptron Navie Bayses, J48	Weka	IMDB	81%
[22]	Multilinear Regression SVM	Weka	Box-office mojo	56.25%
[23]	logistic regression	Weka	Box-office mojo	77.1%

2.4 Neural Network Used in the following Research's

Alexander Jager [24] makes movie success prediction by allowing analyst to choose movies supported by both benchmark datasets i.e. IMDB and Twitter sentiment analysis. Once movie chooses, then they predict rating for that particular movie using NEURAL NETWORK, which takes only two parameters for predicting rating such as (i) ratings of crew members and (ii) actors of relevant movies. They used the following two tools for implementation: VISIONE and

KNIME. VISIONE (Visual social network) is a tool which is used to perform analysis and visualize graphically structures of social networks (Twitter and Facebook dataset etc). The second tool is KNIME (Konstanz Information Miner), it contains various data mining techniques and ability to export trained models in the PMML (Predictive Model Markup Language) format. MYSQL database is use to store data as a data set. Tweets by the users about the movie and IMBD data is used to create a well structured data set. Plain text data is extracted from the IMDB where users give their opinion and rating about the movies. Each movie studio have a team of analyst. They read people opinion and interest and recommend the movies industry to manage their resources accordingly. The model used to compute the rating of movies after first week of movie release. At the end they compare their results witch actual rating given by the users after release of first weekend of the movie. They get better results because of training the neural network on their data set.

Brett Kuprel. [25] movies genre can easily be determined by its poster. There is a lot of work have be done on movies poster. Movies poster can be used to predict movies quality and movies genre etc. A human can easily guess the movie genre just by looking at movies poster but its hard for the computer to judge. In this research the author proposed a method in which a movie quality can be judge by its poster. He applies deep learning algorithm on a given data set of posters. A deep learning is the sub field of machine learning techniques. Its an advance version of machine learning algorithms. Autoencoder approach Is used and identify the genre of the movie just by its poster. IMDB data set is used which is consist of 1000 movies and use multiple posters of the movies. The posters consist of 5800 images. Each image has pixel of 100x100 size. The proposed model is implemented in python and used Theano Package. A neural Network is used in which feed forward and back propagation is applied. There is chance that one movie can present multiple genre in its content. for example a movie back to the future consist of three genre which include adventure, comedy and science fiction. In this way that type of movies will be allocated to multiple genres.

Table 2.4: Summary of Neural Network Used in the Research

Ref.	Techniques Use	Tool Use	Dataset	Accuracy
[24]	Neural Network	Python script, MySQL	IMDB	73%
[25]	Neural Network	Python Theano Package	IMDB	45.65%

2.5 Other Relevant Algorithms that Used to Compute Movies Quality

Rutuja Jadhav [26] Uses movie trailer reviews posted by different users on social media platforms to predict box office revenue. Social media reviews are very useful in predicting outcome of an events. Social media platforms directly involve the end user to know his/her interest and get reviews about

the products. People Freely explore their views with comfort about the movies trailer and also other online stuff. It help both the user and the owner to make their decision accordingly. In this paper, they have collected movie trailer reviews from social sites to predict the movie sales performance, they calculate rating and revenue using sentimental analysis. Movie sales performance is predicted based on reviews given in only English language. A very short reviews or reviews that contain special words are not considered. A data mining tool name as openNLP is used to mine all of the reviews given by the social media users after release of movies trailer and perform sentiment analysis to compute movie sales performance before its release. Online reviews are very helpful in share market. The investor gain some confidence by reading online reviews given by the end users. The investors focus on the people attention for their products and decide whether to invest more or not. They expect a huge ROI when receives positive reviews from the end user. The author perform sentiment analysis on people reviews on movie trailers to get output in two parts, expected rating of movies and expected revenue generated from movies.

Farshad Bakhshandegan Moghaddam [27] Proposed a model that predict movie quality using visual features. They use movielens dataset of 1300 movie trailers that predict the quality of the movie prior to its release. They use the hybrid features including content of the movies to predict not just the popularity but also predict the rapping of the movie. The work is based on the low level feature. They do not consider the user reviews or rapping. They also do not consider the cast of the movies like actor, directors. They only use the actual content that shown in the movie trailer and applied their proposed model on the dataset and predict the movie success. They raise two question one was predicting movie popularity and second was predicting movie average rating. They gives answer to those questions in research. The machine learning algorithms are the best classification algorithms, so they use one of the machine learning algorithm name as Gradients Algorithm in their research which successfully able to predict the movie popularity. The movie popularity is classified in three classes. Popular movie, average popular and un-popular movie. The accuracy rate of predicting movie popularity (first question) was 56% by using visual hybrid features. The accuracy rate of predicting movie average rating (second question) was also give better result by using hybrid features.

Jonas Krauss [28] sentiment analysis plays an important role to predict the success of the movie. Social media platform plays an important role to make things viral. When somethings gets viral peoples took interest to get knowledge about that products. Similarly when movies gets viral on social medial platforms like facebook, twitter then people gets curious about that movie and wish to watch it. Each movie industry have a full time team of social media analyst. They carefully follow the people reviews and work accordingly. So They focus on the social networks and gather the review of different users. They use IMDB data set where people give their reviews and rapping about the movie. On the basis of these discussion they predict whether the movie will be nominated in Oscar award and also predict the box office success revenue. Before two moth of the ceremony of Oscar award they predict which movie will be nominated in Oscar award and which move will receive Oscar. The proposed method can successfully predict nine out of ten ratio that which movie will receive Oscar in this time. They choose two database namely IMBD data set and Box office

Mojo data set. This model can easily be implemented on other online businesses for example in stock market analysis etc.

Table 2.5: Summary of Other Relevant Algorithms

Ref.	Techniques Use	Tool Use	Dataset	Accuracy
[26]	ARSA model Java	OpenNLP MySQL	Sentiment Analysis	73%
[27]	Gradients Algorithm	Python	Movielens	56%
[28]	Web mining	Condor TeCFlow	IMBD Box office Mojo	85%

Chapter 3

Methodology

In this research we have used Machine learning classification algorithms to predict movies quality. Machine learning algorithms are used to classify given set of images or data into the same categories of classes. Machine learning algorithms work exactly like a human brain works. Our brain looks at the image and learn that image. Next time when the brain look at the same image it compares that image with a group of other available images and perform classification to the best fittest group. Similarly, machine learning algorithms are trained by applying some classification algorithms on a given training data set. When algorithm learned by training, then it can easily classify the input by comparing it to the nearest possible class. It's takes some time to train a machine but when it's trained, then it can classify million of images without any fatigue while a human brain cannot do that. In our proposed model, Movie quality is classified in four different categories by using many of the machine learning classification algorithms. A threshold value will be set in the form movies rating given by the users on the IMDB website. These values have a range of 1-10 movies rating, which is shown in the table 3.1. According to that threshold value a movie will be classified as a flop, average, good and Blockbuster. In our proposed work we have used IMDB data set that we get from Kaggle, a big data set repository platform freely available to public. Our data set consist of 1000 Hollywood movies that are released in between the years of 2012 to 2016. The data set contains multiple genres of movies. It contains different attribute like Metascore, Revenue, Votes, Rating, Runtime, Year of release, Actors, Directors, Description, Genre and Title. In our proposed model the main attributes that are used to predict movies quality are Metescore, Rating, Revenue, Runtime, Votes etc. The rating attributes act as label by which we use to classify our movies into different classes. The IMDB is one of the largest movie website where people give rating and reviews about the movies. IMDB data set freely available on the web which contains information about movies. The rating about the movie's performance given by the viewers on IMDB website is very useful for the movie studio and also the viewers. It helps the movie studio to manage their resources accordingly while helping the viewers to choose a best ratted movie to watch it without consuming time in searching. It is one of the most popular sources of movie data features such

as title, actor, director, writer, genre, movie news, movie reviews, trailer, votes, show times and release date etc.

In our proposed model we have used hybrid decision features [29] to compute movies quality. After extracting features from the data set, features scaling techniques such as standardization and Min-Max scaling will apply to normalize and standardize the particular numeric attribute. After that, a well-defined Machine learning algorithms will apply to predict movies quality using extracted hybrid features. After that evaluate each feature and also automate the whole process. In this work, we will label prediction results into four classes based on developed threshold value in the form of movies rating given in the table 3.1. The criteria set for classifying different movies in different classes starting from Flop to Hit is as follow. These classes are made by the movies rating attribute which is used as label or class attribute.

Table 3.1: Class of Movies Consideration

Range of Rating	Class
0.0-2.9	Flop
3.0-4.9	Average
5.0-6.9	Good
7.0-10.0	Blockbuster

3.1 Implementation steps

3.1.1 Selection of Movies Attributes

The first step is to extract movies attributes from the data set such as Metascore, Revenue, Votes, Rating, Runtime, Year, Actors, Directors, Description, Genre and title. In movies quality decision, We have used only those attributes who have numeric values. So we have used Metascore, Revenue, Votes, Rating, and Runtime attributes to predict the movies quality. We have used all these feature to compute which algorithm gives us the best accuracy score among all of the others. We have categories the movies in four different classes shown in table 3.1 on the basis of rating attributes which was used as label or class attribute.

3.1.2 Scaling of extracted features using Training data set

After extraction of attributes, features scaling techniques will apply to scale the attributes. We have used 80% of our data set to training the model. There are different scaling techniques are available such as standardization and Min-Max scaling to normalize and standardize the particular numeric attribute. In our proposed work we have used Min-Max Scaling Techniques because this techniques scale our data more accurately and efficiently as compare to other scaling techniques.

3.1.3 Applying classification algorithm

After Scaling process, classification algorithm will apply to compute the movie quality using those extracted attributes. Its output is merely a decision regarding the class to which a particular movie belongs. In this work, we have applied four machine learning algorithms to predict movies quality. Predication results will be label into four classes such as flop, average, good and block buster.

3.1.4 Developing threshold value

There are many movies that released every year. Each movie have their own sorts of popularity. On the basis of this popularity movies get rating by the user. These rating categories the movies in different classes. A movie can be a flop, average, good or block buster on the basis of these ratings given by the viewers. So we have set a threshold values that are given in the table 3.1 in the form of rating attribute which used as label in our proposed model.

3.1.5 Testing classification results with testing data set

Before evaluating classification results with developed threshold value, these results will test with testing data set. We have used 20% of our data set in our testing phase. since our data set consist of 1000 movies so we have used 200 movies in testing phase to test our proposed model.

3.1.6 Evaluating calculated results with developed threshold value

Once the classification results successfully tested, now it will evaluate with developed threshold value. If the results lies in between 1-10 rating as we have set rating attribute as label, so these specific movies will be allocated to respective class. At the end we will find our that how many movies are classified as flop, average, good and blockbuster, We will also find out that which algorithm gives us the best accuracy rate.

3.1.7 Terminate the algorithm

Our proposed model will gives the accuracy score of each classifier used in the implementation and the algorithm will be terminated.

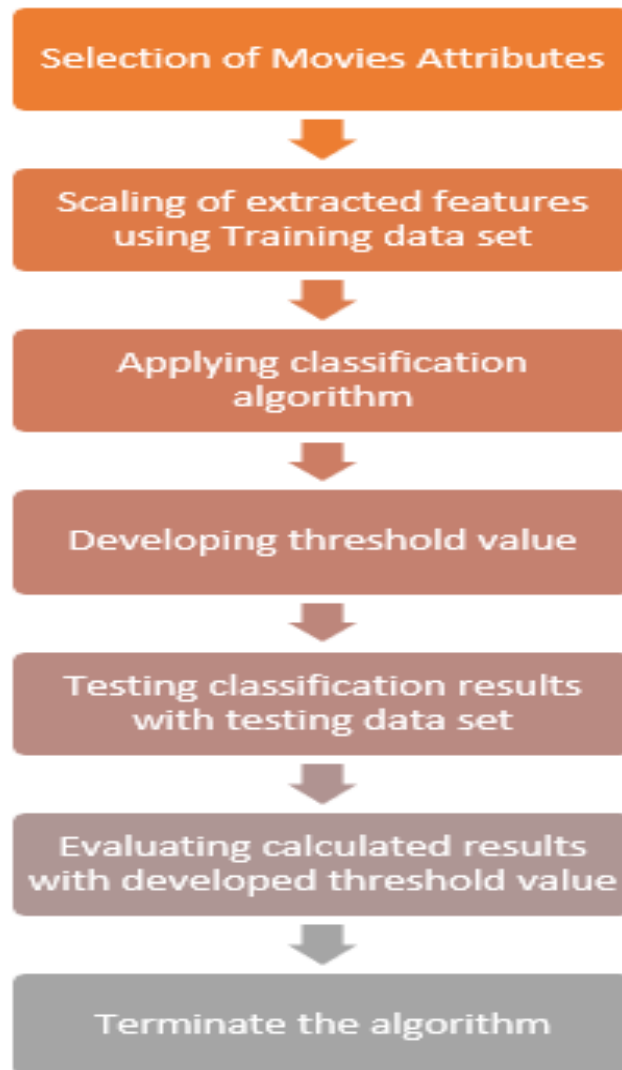


Figure 3.1: Implementation Steps

Chapter 4

Implementation and Results

The proposed model is implemented in anaconda and pycharm in which Python language is used for implementing machine learning algorithms. The model extract movies attributes from a given data set and then it preprocess the data to clean the data for getting good and high quality results. After preprocessing the data we have used a feature scaling process to scale the features and select only those attributes that plays main role in our proposed method. The attributes that do not have any impact in our proposed model will be eliminated. After Scaling we have used Different machine learning algorithms to classify the movies in four different categories, i.e. Flop, Average, Good and Blockbuster. We have set thresh-hold values that are illustrated in 3.1 to classify all those movies in four different categories. We have used four state of the art highly used and impacted machine learning algorithms named as Linear Regression, Decision Tree, Random Forest and Naive Bayes algorithms. We find out that which algorithms gives us the best accuracy score among all of the others.

4.1 System Architecture

In this research we have used different machine learning algorithms to calculate movies quality in term of flop, average, good and Blockbuster and to find out which algorithms gives us the higher accuracy score. We have used anaconda in which python language is used to implement all these machine learning classification algorithms. We have used a very simple setup to implement our model. Our system configuration is Intel(R) Core(TM) i5-6200 CPU@2.30GHz 2.40GHz and 4GB RAM and 64 Bit Operating System. That type of system can easily be available to any of the users because its very common and very affordable. We have used IMDB Movie data set that we get from kaggle, freely available on internet that consist of 1000 Hollywood movies. All these movies were released in between 2012 to 2016 in which most of the movies were released in 2016. Our focus was to get the data set that contains latest released movies, in which we got succeeded. The data set consist of different movies attributes, for example Metascore, Revenue, Votes, Rating, Runtime,

Year, Actors, Directors, Description, Genre and title. The main attributes that play an important role in the classification were rating, Metascore, revenue, runtime and votes. We have feature scaling techniques to scale the feature and used only those features that have a huge impact in movies quality prediction. We then preprocess the data to get the higher quality and better results.

Our proposed model works in following steps.

- Data Extraction
- Data Preprocessing
- Feature Selection
- Apply machine learning algorithms
- Results of machine learning algorithms

4.1.1 Data Extraction

In this phase we have collected our movies related data from online data set repository called Kaggle. Our focus was to search the latest movies data set in which we get succeeded. All movies were English movies. Our data set contains movies that were released from 2012 to 2016 in which maximum movies were released in 2016. Our data set contains about 1000 Hollywood movies.

4.1.2 Feature Selection

Feature selection phase is used when data set contains some irrelevant attributes or those attributes who contributes less in experiments. In our work we remove those attributes that do not contribute in our proposed model. Feature selection can be implemented by many methods, but in our work we have used the information gain approach. By using this approach we have successfully performed feature selection and identify attributes either they have any impact in in our analysis or not.

4.1.3 Data Preprocessing

In this phase we have prepared our data set for classification according to our need. The data set that we obtain contain some noise and missing values. That type of data needs to be cleaned before processing it, that phase is called data preprocessing phase. We have removed some of the missing values and in some cases we have applied mean and average strategy to fill these blank spots. The feature of the data set are Metascore, Revenue, Votes, Rating, Runtime, Year, Actors, Directors, Description, Genre and title. The Rating attribute was used as class attribute or as label.

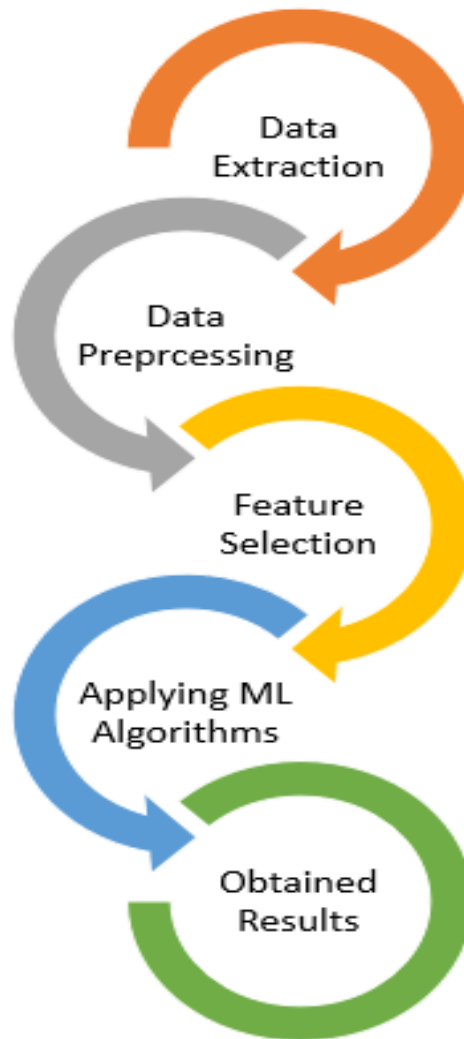


Figure 4.1: System Architecture

4.1.4 Apply machine learning algorithms

The proposed model implements in Anaconda in Jupyter, in which Python language is used for implementing the machine learning algorithms to predict the movies quality whether it is flop, average, good or blockbuster. In our proposed work we have applied four machine learning algorithms to find out movies quality and also find out which classifier gives us the highest accuracy score. We have applied Linear Regression, Decision Tree, Random Forest, and Naive Bayes. Out of these machine learning algorithms Decision Tree gives the best accuracy score among the other classifiers. While the Naive Bayes gives us the least accuracy score among the other algorithm.

Table 4.1: List of Classifiers Used in Proposed Model

Name of classification algorithms
Linear Regression
Decision Tree
Random Forest
Naive Bayes

4.2 Machine Learning Classification Algorithm

4.2.1 Linear Regression

It is one of the most popular machine learning algorithm. It is use to predict the results in given range. That's why We have used this algorithm to predict the movies quality in four different categories. It is use to get the categorical output. In our proposed work the linear regression gives us the accuracy score of 82%.

4.2.2 Decision Tree

Decision Tree is also known as J48. It gives the output in form of tree like structure. In decision Tree the most likely data is allocated to the same class. Since our work is to classify the movies quality in four different categories that's why we have used this algorithms in our research. In our proposed work the decision tree gives us the accuracy score of 91.4%.

4.2.3 Random Forest

Random forest also work same like decision tree because it also form a tree like structure but in a different way. Random forest use in both classification and regression problems. Since our work is to classify the movies quality in four different categories that's why we have used this algorithms in our research. In our proposed work the random forest gives us the accuracy score of 82%.

4.2.4 Naive Bayes

Naive Bayes algorithm used to perform classification. It is very popular for its simplicity. we have used this algorithm because of its performance and the prediction speed. It is very helpful in small data set. Training of this model is very easy. This algorithm assume feature as independent but in our model all the feature are dependent to each other that's why it's accuracy score was lower than other algorithms. In our proposed work the Naive Bayes gives us the accuracy score of 79.56%.

4.3 Libraries

The proposed model implemented in Anaconda in which Python language is used. We have used following libraries in our proposed model.

- Pandas
- Matplotlib
- Numpy
- sklearn

4.4 Data set

In this research we have used real world IMDB movies data set. We obtain this data set from kaggle, freely available to the public. Our data set consisted of 1000 movies which includes all of the concerns feature that are used in the movies. It contains different attribute like Metascore, Revenue, Votes, Rating, Runtime, Year of release, Actors, Directors, Description, Genre and Title. We have used feature scaling techniques to scale the features given in the data set. In our proposed model the main attributes that were used to predict movies quality are Metescore, Rating, Revenue,Runtime, Votes etc. We have used rating attribute as a label or class. Using rating attribute we have classified that how many movies are belong to different classes. These classes are formed as rating of 1-10 which is shown in the table 3.1 There was some missing value and noise in the data set so we first clean the data in data preprocessing phase.

We have used 80% of the data in our data set in the training phase to train the algorithms effectively and efficiently. The remaining 20% of the data was used in a testing phase to test the accuracy score of the proposed model. Since our data set consist of 1000 movies, so 800 movies were used to train the model and the remaining 200 movies were used to test the model. We have split the data into training and testing by using the python library sklearn. We have used stratified shuffle function to shuffle the data so that we can maintain transparency in splitting the data set into training and testing data.

4.5 Results of our proposed model

In our proposed model we have used four state of the art very common and useful machine learning algorithms named as Linear Regression, Decision Tree, Random Forest and Naive Bayes. These algorithms were used to classify movies quality in four different categories as a flop, average, good and Blockbuster movies. We have used rating attributes as a label to compute that how many movies belong to a particular class. We have obtained our IMDB data set from kaggle, freely available to public. Our data set was consisted of 1000 Hollywood movies so we have used 800 movies in the training phase to train the proposed model. We have used remaining 200 movies in the testing

phase to test our model. Our experiment shows that there were a very less number of movies were selected as flop movies. So We have used SMOTE Technique by which we can get about 54 movies as flop movies and the same as with Block Buster. Our experiment results show that Decision Tree algorithms give us the highest accuracy score among the other algorithms that were used. We have used 10 fold cross validation in which we get random accuracy score for each algorithm. The accuracy Score of decision tree was 91.4% and the accuracy score of linear regression and the random forest algorithm was almost the same which was computed as 82%. While the accuracy score of Naive Bayes algorithm was 79.56%.

We have used the matplotlib library to compute the results in a shape of the graph. We have used different attribute with the combination of rating attribute which was used as a label attribute. Using this combination we have chosen Rating attribute as x-axis and the all of the other attributes as the y-axis. For example, we have chosen Revenue as the y-axis and Rating as x-axis so when the revenue increases, then it's quite possible that movie's rating will also be increase and vice versa. In this way, movies could either be a flop, average, good or Blockbuster. We have used feature scaling technique to scale the features that participate effectively in our proposed model. These features include Metascore, runtime, votes, revenue, Rating etc. We have plotted the graph of all of those features that are obtained by feature scaling techniques with the combination of label attribute. So we have used all of these concern attributes with the combination of rating attributes to find out the results in the form of graph. These results are shown in the following figures.

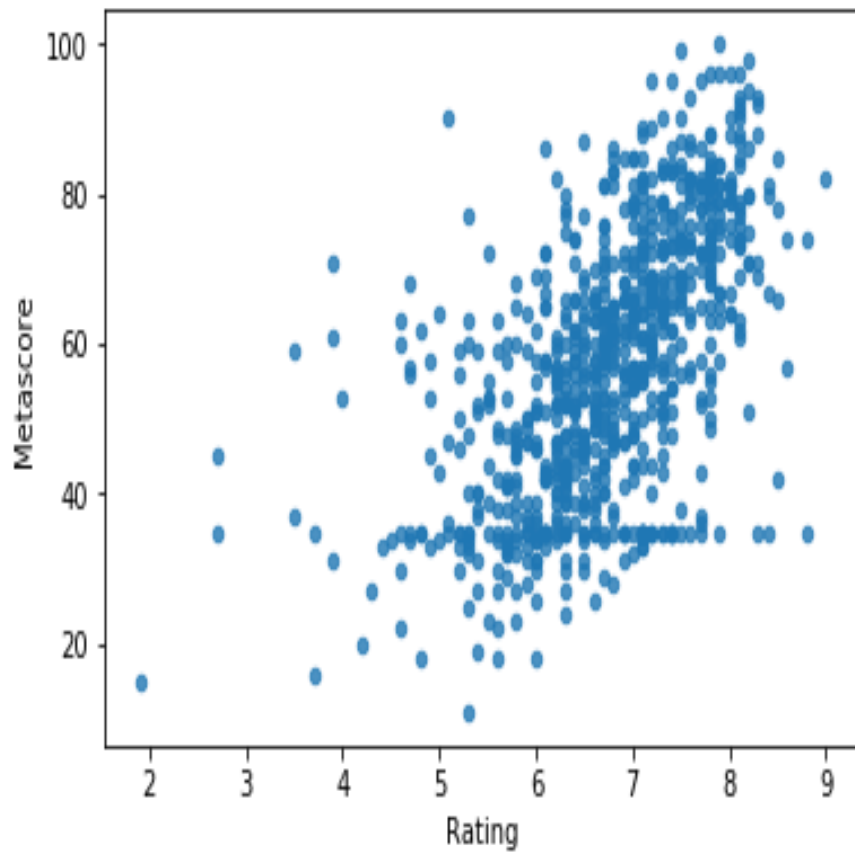


Figure 4.2: Effect of Metascore on Rating

Metascore is actually people reviews on movies. The higher the metascore the movie will be consider as more popular. As metascore increase the movies rating will also increase and vice versa. So the movies quality depend on metascore of the movies, when we compare it with rating attribute which work as label attribute in our proposed work as shown in [4.2](#)

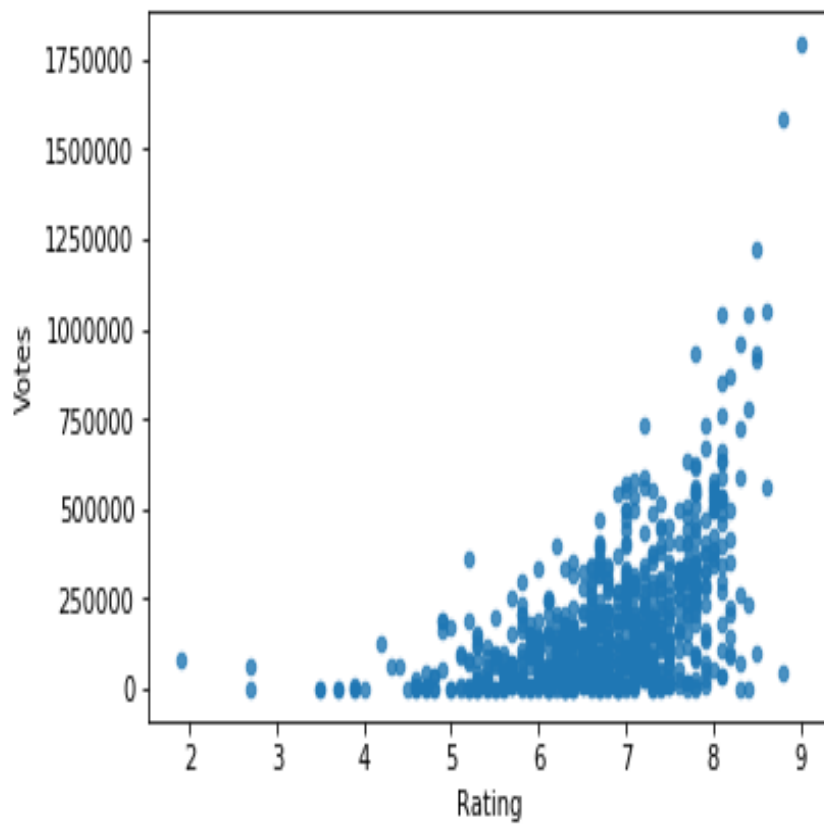


Figure 4.3: Effect of Votes on Rating

Votes are actually people interest in a movies. The higher the Votes the movie will be consider as more popular. As Votes increase the movies rating will also increase and vice versa. So the movies quality depend on votes of the movies, when we compare it with rating attribute which work as label attribute in our proposed work as shown in [4.3](#)

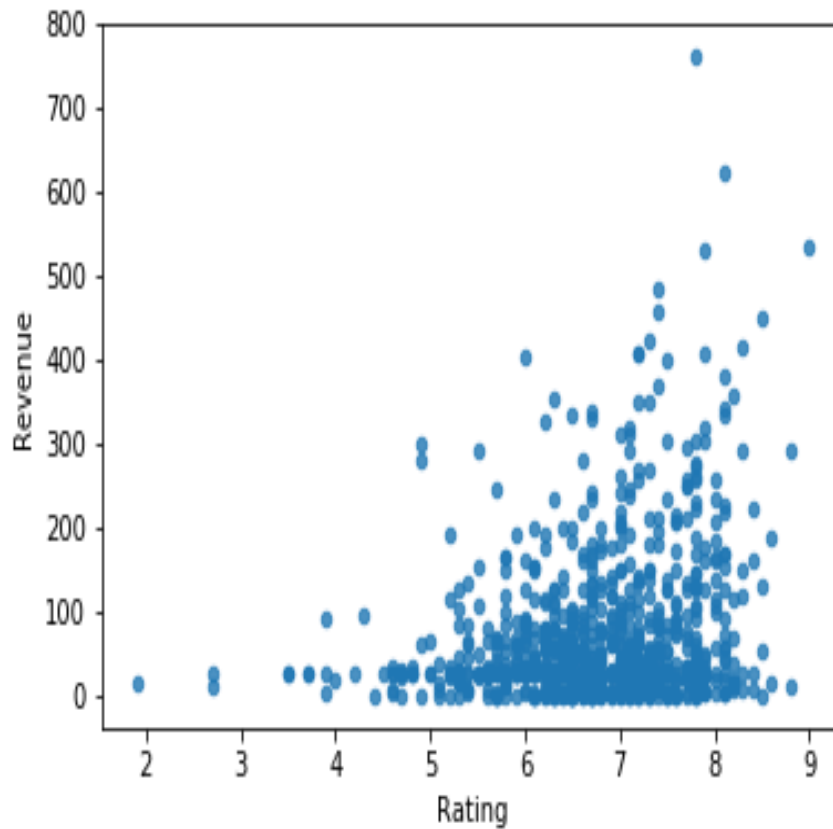


Figure 4.4: Effect of Revenue Generated by Movie on Rating

Revenue is generated from the viewers. The more viewers will watch the movie the more revenue will be generated. Revenue shows that how much movie is popular. As revenue increase the movies rating will also increase and vice versa. So the movies quality depend on revenue of the movies, when we compare it with rating attribute which work as label attribute in our proposed work as shown in [4.4](#)

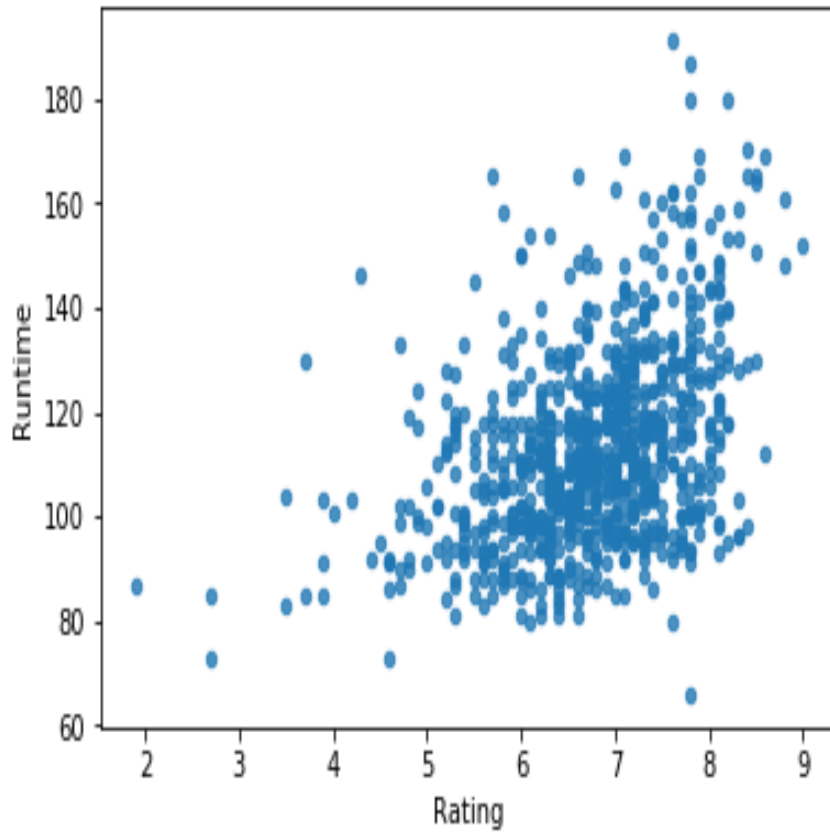


Figure 4.5: Effect of Runtime on Rating

Runtime shows the actual runtime of the movie in which a particular movie is screened. The more viewers will watch the movie the more runtime will be counted. Runtime shows that how much movie is played. As runtime increase the movies rating will also increase and vice versa. So the movies quality depend on runtime of the movies, when we compare it with rating attribute which work as label attribute in our proposed work as shown in [4.5](#)

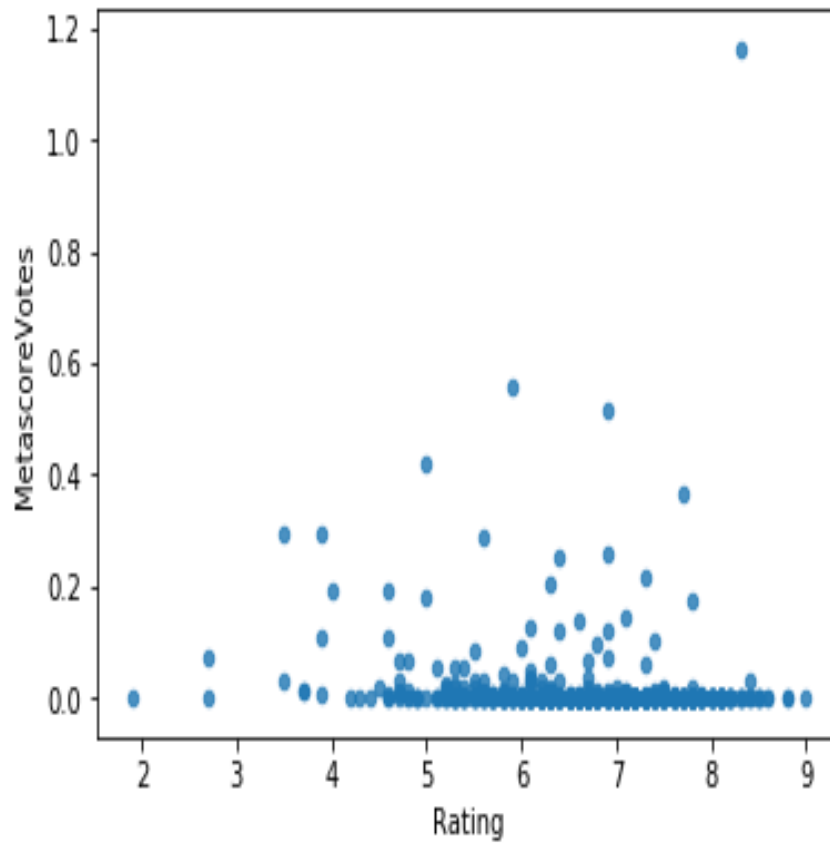


Figure 4.6: Effect of Combination of Metascore & Votes on Rating

As an experiment we combine the metascore and votes and consider it as one attribute. Then we compare it with rating attribute and draw the graph. It shows that how much a movie attract the viewers. So the movies quality depend on metascore votes, when we compare it with rating attribute which work as label attribute in our proposed work [4.6](#)

Chapter 5

Conclusions

In this last chapter, we will briefly describe an overall work of the research carried out as a proposed method and findings and will also propose some future work.

Movies quality prediction is now very common and very important part of the movies studio as well as the viewers. Each movie is produced by the millions of dollars of investment. There is no chance to get it flop because it makes loss millions of dollars to the movie industry. There is a huge team and a very hard work behind each movie to make it a successful and super hit at the box office. The movies quality prediction becomes more common when Netflix announces Netflix prize in 2006. They announce that whoever increase 10% accuracy score in the existing movies quality prediction model will win a huge prize. The movie industry uses past and present data and perform sentiment analysis to predict their quality of upcoming movies. They also use well known machine learning and deep learning algorithms to know where their movie will stand in the future in term of popularity and revenue generation, when it will be released. All movie industry release trailer of each new movie before its release to know people's interest and then make some required minor changes.

In this research we have used real world IMDB movies data set. We obtained our data set from a well known data set repository called as kaggle, freely have access to public. Our data set consist of 1000 movies which includes all of the concerns feature that are used in the movies. We have used 800 movies in the training phase to train our model and the remaining 200 movies were used to test our proposed model results. It contains different attribute like Metascore, Revenue, Votes, Rating, Runtime, Year of release, Actors, Directors, Description, Genre and Title. We have used feature scaling techniques to scale our given data set attributes. In our proposed model the main attributes that were used to predict movies quality are Metescore, Rating, Revenue, Runtime, Votes etc. There was some missing value and noise in the data set so we first clean the data in data preprocessing phase. In data preprocessing phase, we have clean the data, we have removed some of the missing values and we also use mean, average and standard deviation to add some missing values where possible.

We have used four state of the art and highly used machine learning algorithms to predict movies quality in the form of flop, average, good or Blockbuster. We have implemented our proposed model in anaconda in jupyter using python language. We have also used different libraries in implementation. We have used a very simple system to implement our proposed model. Our system configuration was Intel(R) Core(TM) i5-6200 CPU @2.30GHz 2.40GHz and 4GB RAM and 64 Bit Operating System. This system can easily be accessible and affordable for any of the users. We have used four machine learning algorithm from which decision tree gives the highest accuracy rate from all of the other classifier. We have used 10 fold cross validation in which we get random accuracy score for each algorithm. The accuracy Score of decision tree was 91.4% and the accuracy score of linear regression and the random forest algorithm was almost the same which was computed as 82%. While the accuracy score of the Naive Bayes algorithm was 79.56%.

5.1 Future Works

The proposed model is good for our current system configuration using all of the same rules and techniques But there is room for future work. It can be implemented using other machine learning algorithm and with heavy system configuration in a more efficient way. The experiment can be used in other commercial and marketing product to compute the quality of the product by using required data set. So this model can also be implemented for a marketing purpose. This model can be implemented at a larger scale using heavy system configuration and the related data set.

References

- [1] Garima Verma and Hemraj Verma. Predicting bollywood movies success using machine learning technique. pages 102–105, 02 2019. Cited on p. 1.
- [2] Mladen Marovic, Marko Mihokovic, Mladen Miksa, Sinisa Pribil, and Alan Tus. Automatic movie ratings prediction using machine learning. pages 1640–1645, 01 2011. Cited on p. 1.
- [3] L. T. Ponnampalani, S. Deepak Punyasamudram, S. N. Nallagulla, and S. Yellamati. Movie recommender system using item based collaborative filtering technique. In *2016 International Conference on Emerging Trends in Engineering, Technology and Science (ICETETS)*, pages 1–5, 2016. Cited on p. 2.
- [4] Griffin Scott. Determinants of box office performance: Return of the regressions. pages 71–83, 01 2019. Cited on p. 5.
- [5] Hao Wang, Naiyan Wang, and Dit-Yan Yeung. Collaborative deep learning for recommender systems. 08 2015. Cited on p. 7.
- [6] Andrei Oghina, Mathias Breuss, Manos Tsagkias, and Maarten Rijke. Predicting imdb movie ratings using social media. pages 503–507, 04 2012. Cited on p. 7.
- [7] Paul Covington, Jay Adams, and Emre Sargin. Deep neural networks for youtube recommendations. pages 191–198, 09 2016. Cited on p. 9.
- [8] Cyril Garcia and Luca Rona. The netflix challenge. 05 2018. Cited on p. 9.
- [9] Warda Bristi, Zakia Zaman, and Nishat Sultana. Predicting imdb rating of movies by machine learning techniques. pages 1–5, 07 2019. Cited on pp. 9 and 11.
- [10] Nahid Quader, Md Gani, and Dipankar Chaki. Performance evaluation of seven machine learning classification techniques for movie box office success prediction. pages 1–6, 02 2018. Cited on pp. 10 and 11.
- [11] Rui Yao and Jianhua Chen. Predicting movie sales revenue using online reviews. pages 396–401, 12 2013. Cited on pp. 10 and 11.
- [12] P. Nagamma, H.R. Pruthvi, K.K. Nisha, and N.H. Shwetha. An improved sentiment analysis of online movie reviews based on clustering for box-office prediction. *International Conference on Computing, Communication and Automation, ICCCA 2015*, pages 933–937, 07 2015. Cited on pp. 10 and 11.
- [13] Mladen Marovic, Marko Mihokovic, Mladen Miksa, Sinisa Pribil, and Alan Tus. Automatic movie ratings prediction using machine learning. pages 1640–1645, 01 2011. Cited on pp. 11 and 13.

- [14] P. Li and S. Yamada. A movie recommender system based on inductive learning. pages 318 – 323 vol.1, 01 2015. Cited on pp. 12 and 13.
- [15] HengSong Tan and HongWu Ye. A collaborative filtering recommendation algorithm based on item classification. pages 694–697, 05 2017. Cited on pp. 12 and 13.
- [16] Rahul Katarya. Movie recommender system with metaheuristic artificial bee. *Neural Computing and Applications*, 30, 01 2018. Cited on pp. 12 and 13.
- [17] Zan Wang, Xue Yu, Nan Feng, and Zhenhua Wang. An improved collaborative movie recommendation system using computational intelligence. *Journal of Visual Languages Computing*, 25, 10 2014. Cited on pp. 13 and 14.
- [18] Mennatallah El-Assady, Daniel Hafner, Michael Blumenschein, Alexander Jäger, Wolfgang Jentner, Christian Rohrdantz, Fabian Fischer, Svenja Simon, Tobias Schreck, and Daniel Keim. Visual analytics for the prediction of movie rating and box office performance. 01 2013. Cited on pp. 14 and 16.
- [19] Robert Krueger, Harald Bosch, Dennis Thom, Edwin Puettmann, Qi Han, Steffen Koch, Florian Heimerl, and Thomas Ertl. Prolix-visual prediction analysis for box office success. 01 2013. Cited on pp. 14 and 16.
- [20] Dr.R.R.Sedamkar Ms.Mary Margarat Valentine, Ms. Veena Kulkarni. A model for predicting movie’s performance using online rating and revenue. 09 2013. Cited on pp. 14 and 16.
- [21] Muhammad Hassan Latif and Hammad Afzal. Prediction of movies popularity using machine learning techniques. 2016. Cited on pp. 15 and 16.
- [22] Subramaniaswamy V, M. Vaibhav, R. Prasad, and Logesh R. Predicting movie box office success using multiple regression and svm. pages 182–186, 12 2017. Cited on pp. 15 and 16.
- [23] Michael Lash and Kang Zhao. Early predictions of movie success: The who, what, and when of profitability. *Journal of Management Information Systems*, 33, 06 2015. Cited on pp. 15 and 16.
- [24] Mennatallah el Assady, D. Hafner, M. Hund, Alexander Jäger, Wolfgang Jentner, Christian Rohrdantz, F. Fischer, S. Simon, T. Schreck, and D. Keim. Moovis – a visual analytics tool for the prediction of movie viewer ratings and boxoffice. 2013. Cited on pp. 16 and 17.
- [25] Brett Kuprel. Judging a movie by its poster using deep learning. 2013. Cited on p. 17.
- [26] Rutuja Jadhav Pooja Mhaske pooja Deshmukh, Shraddha Bhalerao. Movie sales performance prediction. pages 74–79, 03 2015. Cited on pp. 17 and 19.
- [27] Farshad Bakhshandegan Moghaddam, Mehdi Elahi, Reza Hosseini, Christoph Trattner, and Marko Tkalčič. Predicting movie popularity and ratings with visual features. 06 2019. Cited on pp. 18 and 19.
- [28] Jonas Krauss, S. Nann, Daniel Simon, Kai Fischbach, and Peter Gloor. Predicting movie success and academy awards through sentiment and social network analysis. 06 2018. Cited on pp. 18 and 19.

- [29] Xinxi Wang and Ye Wang. Improving content-based and hybrid music recommendation using deep learning. pages 627–636, 11 2014. Cited on p. [22](#).

Thesis

ORIGINALITY REPORT

12%

SIMILARITY INDEX

9%

INTERNET SOURCES

5%

PUBLICATIONS

8%

STUDENT PAPERS

PRIMARY SOURCES

1

Submitted to University of Limerick

Student Paper

2%

2

docplayer.net

Internet Source

1%

3

Submitted to Higher Education Commission
Pakistan

Student Paper

1%

4

www.ijitee.org

Internet Source

<1%

5

"Intelligent Data Communication Technologies
and Internet of Things", Springer Science and
Business Media LLC, 2020

Publication

<1%

6

eprints.soton.ac.uk

Internet Source

<1%

7

"International Conference on Innovative
Computing and Communications", Springer
Science and Business Media LLC, 2021

Publication

<1%

8	krishikosh.egranth.ac.in Internet Source	<1%
9	Submitted to Coventry University Student Paper	<1%
10	hal.archives-ouvertes.fr Internet Source	<1%
11	d-nb.info Internet Source	<1%
12	Bogdan Walek, Vladimir Fojtik. "A hybrid recommender system for recommending relevant movies using an expert system", Expert Systems with Applications, 2020 Publication	<1%
13	pnrsolution.org Internet Source	<1%
14	Submitted to IIT Delhi Student Paper	<1%
15	Submitted to Mancosa Student Paper	<1%
16	bib.dbvis.de Internet Source	<1%
17	"Intelligent Information and Database Systems", Springer Science and Business Media LLC, 2020 Publication	<1%

18

Submitted to Imperial College of Science,
Technology and Medicine

Student Paper

<1%

19

Bagher Rahimpour Cami, Hamid Hassanpour,
Hoda Mashayekhi. "A content-based movie
recommender system based on temporal user
preferences", 2017 3rd Iranian Conference on
Intelligent Systems and Signal Processing
(ICSPIS), 2017

Publication

<1%

20

iosrjen.org

Internet Source

<1%

21

www.mdpi.com

Internet Source

<1%

22

"Inventive Communication and Computational
Technologies", Springer Science and Business
Media LLC, 2021

Publication

<1%

23

"Advances in Information Communication
Technology and Computing", Springer Science
and Business Media LLC, 2021

Publication

<1%

24

gottwurfelt.com

Internet Source

<1%

25

"Emerging Research in Computing, Information,

<1%

Communication and Applications", Springer
Science and Business Media LLC, 2018

Publication

26

"Advances in Computing and Data Sciences",
Springer Science and Business Media LLC,
2019

Publication

<1%

27

dblp.dagstuhl.de

Internet Source

<1%

28

ieevis.org

Internet Source

<1%

29

Farshad B. Moghaddam, Mehdi Elahi, Reza
Hosseini, Christoph Trattner, Marko Tkalcić.
"Predicting Movie Popularity and Ratings with
Visual Features", 2019 14th International
Workshop on Semantic and Social Media
Adaptation and Personalization (SMAP), 2019

Publication

<1%

30

shodhganga.inflibnet.ac.in

Internet Source

<1%

31

www.wanghao.in

Internet Source

<1%

32

"Advances in Intelligent Data Analysis XVI",
Springer Science and Business Media LLC,
2017

Publication

<1%

33

mafiadoc.com

Internet Source

<1%

34

Denis Baylor, Levent Koc, Chiu Yuen Koo, Lukasz Lew et al. "TFX", Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '17, 2017

Publication

<1%

35

"Recommender System with Machine Learning and Artificial Intelligence", Wiley, 2020

Publication

<1%

36

Submitted to National College of Ireland

Student Paper

<1%

37

Submitted to City University of Hong Kong

Student Paper

<1%

38

Submitted to Universiteit van Amsterdam

Student Paper

<1%

39

Wang Zhou, Jianping Li, Malu Zhang, Yazhou Wang, Fadia Shah. "Deep Learning Modeling for top-N Recommendation with Interests Exploring", IEEE Access, 2018

Publication

<1%

40

researchr.org

Internet Source

<1%

41

Submitted to Heriot-Watt University

Student Paper

<1%

42

Chen Dong, Ziqi Xiong, Ximeng Liu, Yin Ye, Yang Yang, Wenzhong Guo. "Dual-Search Artificial Bee Colony algorithm for Engineering Optimization", IEEE Access, 2019

Publication

<1%

43

koreascience.or.kr

Internet Source

<1%

44

Zahabiya Mhowwala, A. Razia, Sujala D.. "Movie Rating Prediction using Ensemble Learning Algorithms", International Journal of Advanced Computer Science and Applications, 2020

Publication

<1%

45

www.ijedr.org

Internet Source

<1%

46

Ashutosh Kanitkar. "Bollywood Movie Success Prediction using Machine Learning Algorithms", 2018 3rd International Conference on Circuits, Control, Communication and Computing (I4C), 2018

Publication

<1%

47

www.ir.nctu.edu.tw

Internet Source

<1%

"Information Security and Privacy", Springer

48	Science and Business Media LLC, 2018 Publication	<1%
49	www.xlstat.com Internet Source	<1%
50	www.cvast.tuwien.ac.at Internet Source	<1%
51	www.doc4net.com Internet Source	<1%
52	trap.ncirl.ie Internet Source	<1%
53	baadalsg.inflibnet.ac.in Internet Source	<1%
54	Xing, H.J.. "An adaptive fuzzy c-means clustering-based mixtures of experts model for unlabeled data classification", Neurocomputing, 200801 Publication	<1%
55	downloads.hindawi.com Internet Source	<1%
56	www.ijser.org Internet Source	<1%
57	Syed Muhammad Raza Abidi, Yonglin Xu, Jianyue Ni, Xiangmeng Wang, Wu Zhang. "Popularity prediction of movies: from statistical	<1%

modeling to machine learning techniques",
Multimedia Tools and Applications, 2020

Publication

58

Ankit A. Sinha, S. V. Vamsi Krishna, Rajashree Shedge, Avi Sinha. "Movie production investment decision system", 2017 International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS), 2017

Publication

59

elib.uni-stuttgart.de

Internet Source

60

ieeeprojects.eminents.in

Internet Source

61

tel.archives-ouvertes.fr

Internet Source

62

Maganti Venkatesh, S. Sathyalakshmi. "Smart learning using personalised recommendations in web-based learning systems using artificial bee colony algorithm to improve learning performance", Electronic Government, an International Journal, 2020

Publication

63

"Computation and Communication Technologies", Walter de Gruyter GmbH, 2016

Publication

<1%

<1%

<1%

<1%

<1%

<1%

Exclude quotes Off

Exclude matches Off

Exclude bibliography Off