TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 1

1.1 Introduction

Movie industry is a huge sector for investment but larger business sectors have more complexity and it is hard to choose how to invest. Big investments come with bigger risks. The CEO of Motion Picture Association of America (MPAA) J. Valenti mentioned that ‘No one can tell you how a movie is going to do in the marketplace. Not until the film opens in darkened theatre and sparks fly up between the screen and the audience’. As movie industry is growing too fast day by day, there are now huge amount of data available on the internet, which makes it an interesting field for data analysis. Predicting a movie success is a very complex task to do. The definition of a movie success is relative, some movies are called successful based on its worldwide gross income, and some movies may not shines in business part but can be called successful for good critics review and popularity. There are many movies which did not produce good amount of profit during its release time but become famous after few years. In this project we have considered a movie success based on its star cast, the directors and the genre of the film only. The movies made by directors in the past with the actor’s performance and if it was a success or a NOT SUCCESSFUL have been taken into consideration for predicting their next movie in the making.

1.2 Problem Statement

Making a prediction of society’s reaction to a new product in the sense of popularity and adaption rate has become an emerging field of data analysis. The motion picture industry is a multi-billion dollar business. And there is a huge amount of data related to movies is available over the internet and that is why it is an interesting topic for data analysis. Machine learning is a novel approach for analyzing data. Our paper proposes a decision support system for movie investment sector using machine learning techniques. In that case, our system will help investors related with this business to avoid investment risks. The system will predict an approximate success rate of a movie based on its profitability by analyzing historical data from different sources like IMDb, Rotten Tomato, Box Office Mojo and Meta Critic. Using different machine learning algorithms and other techniques the system will predict a movie box office profit based on some features like who are the star cast, director members and the genre of the movie and then process that data for classification.

1.3 Scope

This project will be more helpful in the movie industry. It covers areas which include the pre factors affecting the rating of the movie. The Naive Bayes Classification will be used for classifying the success rate of the movie. It will be of great help for investors to carry out early predictions. Also the accuracy of individual techniques can be measured to decide which model works best in this scenario.

1.4 Applications

This project has its application in the Movie industry. Our mission is to make a model which can help investors to avoid risks and make a right choice of investment. This research will not only help investors but also will be helpful for the whole movie industry. There are many new artists who cannot make a film because no investor is ready to invest for them. Investors have their own reason, not all investor has the courage to invest on a movie of a new director because he/she has no experience to show but they are extremely talented and passionate about film making. Early prediction will help an investor to make choice if he/she wants to invest for new artists. This will be great for new artist in the movie industry. A movie industry contributes a massive amount of money in global economy, everything is connected now in 2017. So if new artists can make movie easily more artist will try to make films, more films will produced day by day and movie industry will contribute more money to global economy. We have made our dataset based on english movies only with pre release features. To predict an upcoming movie only pre-release features will be responsible for prediction. For multiclass prediction several machine learning algorithms are available like Naive Bayes, Support Vector Machine (SVM) and Logistic Regression etc. These classifiers are good enough for binary classification.

CHAPTER 2

2.1 Design

Movie Industry is a multi billion dollar Industry which has been growing at a tremendous rate. Every movie has a huge amount of money riding on its success and hence the probability of success or failure of the movie is an interesting factor to know before the investors can become potential producers for a particular movie. This success can be predicted using various aspects and parameters. The major or key factors which cause the success of a movie are the actors, directors, genre, public interest, rating, etc. The past data of movies is stored and the probability of each factor is calculated. We used a couple of research papers as references to build our project. These papers gave us the required understanding and motivation to complete our project and help in predicting the results. The paper titled "Performance Evaluation of Seven Machine Learning Classification Techniques for Movie Box Office Success Prediction" by Nahid Quader, Md. Osman Gani and Dipankar Chaki. We read their paper and researched about various techniques and models used in Machine Learning and Classification. We watched videos and reviewed codes about various models like Logistic Regression, Support Vector Machine, Random Forest, etc. before settling upon Gaussian Naive Bayes algorithm. This method seems the most viable method for this project.

2.2 Literature Survey

Success of a movie primarily depends on the perspectives how the movie has been justified. In early days, a number of people prioritized gross box office revenue ([1], [2], [3], [4]), initially. Few previous work ([4], [5], [6]), portend gross of a movie depending on stochastic and regression models by using IMDb data. Some of them categorized either success or NOT SUCCESSFUL based on their revenues and apply binary classifications for forecast. The measurement of success of a movie does not solely depend on revenue. Success of movies rely on a numerous issues like actors/actresses, director, time of release, background story etc. Further few people had made a prediction model with some pre-released data which were used as their features [7]. In most of the case, people considered a very few features. As a result, their models work poorly. However, they ignored participation of audiences on whom success of a movie mostly depends. The accuracy of prediction lies on how big the test domain is. A small domain is not a good idea for measurement. Again most of them did not take critics reviews in account. Besides, user's reviews can be biased as a fan of actor/actress may fail to give unbiased opinion. In [11] A. Sivasantoshreddy, P. Kasat, and A. Jain tried to predict a movie box-office pre release prediction using hype analysis. Main logic behind hype analysis is a success of a movie heavily depends on its opening weekend income and also how much hype it gets among people before release based on the past performances of the star caste and the director including the genre of the movie.

2.3 Technique Used

The first phase is data acquisition. Here we choose four data sources IMDb, Rotten Tomato, Box Office Mojo and Meta Critic. Different types of features are extracted from different sources. Second phase is data cleaning. After scrapping data from various sources, we cleaned our data mainly depend on unavailability of some features. After cleaning all data, next phase is data integration and transformation. In third phase we classified some features. Fourth phase is analysis of IMDb and Rotten Tomato reviews. Fifth phase is Result and Analysis, where we applied Naive Bayes Classification on our dataset.

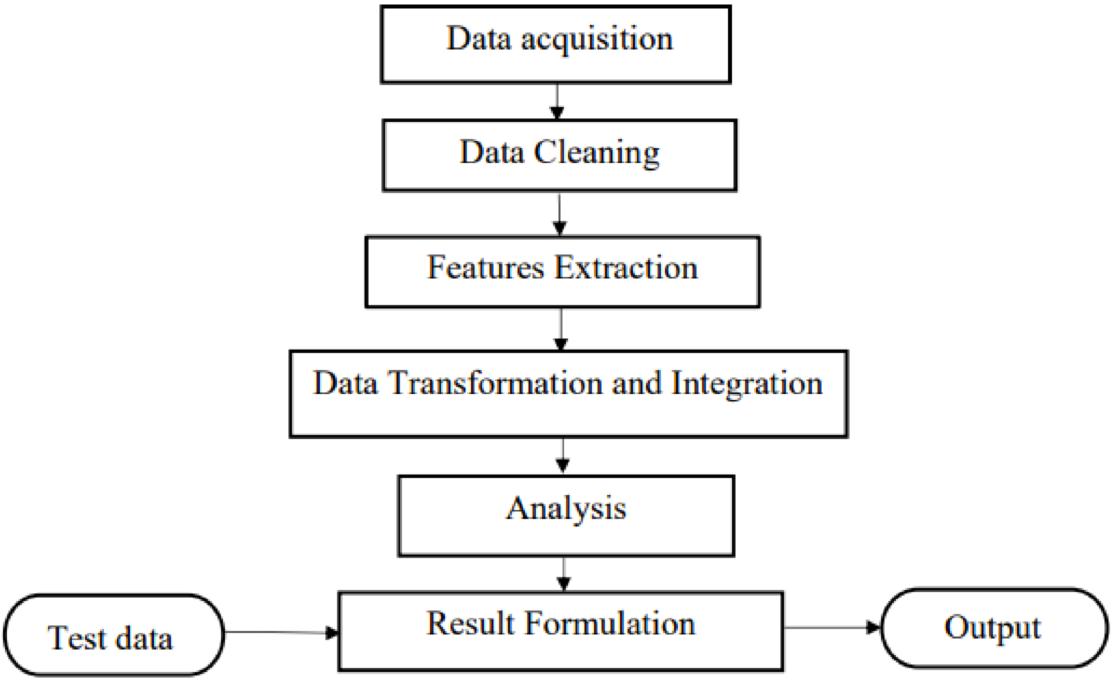


Fig 2.3.1: Research Workflow

Naive Bayes Classification:

Naive Bayes is a simple, yet effective and commonly-used, machine learning classifier. It is a probabilistic classifier that makes classifications using the Maximum A Posteriori decision rule in a Bayesian setting. It can also be represented using a very simple Bayesian network. Naive Bayes classifiers have been especially popular for text classification, and are a traditional solution for problems such as spam detection.

The goal of any probabilistic classifier is, with features x_0 through x_n and classes c_0 through c_k, to determine the probability of the features occurring in each class, and to return the most likely class. Therefore, for each class, we want to be able to calculate P(c_i | x_0, …, x_n).

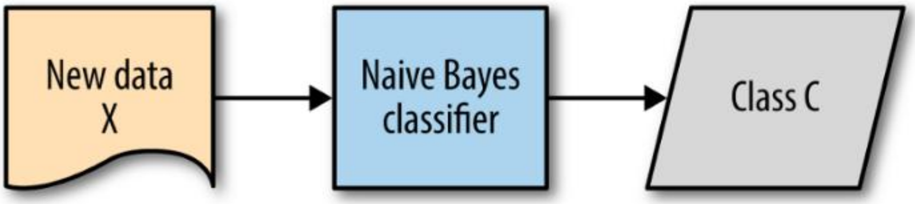
In order to do this, we use Bayes rule. Recall that Bayes rule is the following:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

In the context of classification, you can replace A with a class, c_i, and B with our set of features, x_0 through x_n. Since P(B) serves as normalisation, and we are usually unable to calculate P(x_0, …, x_n), we can simply ignore that term, and instead just state that P(c_i | x_0, …, x_n) ∝ P(x_0, …, x_n | c_i) * P(c_i), where ∝ means "is proportional to". P(c_i) is simple to calculate; it is just the proportion of the data-set that falls in class i. P(x_0, …, x_n | c_i) is more difficult to compute. In order to simplify its computation, we make the assumption that x_0 through x_n are **conditionally independent** given c_i, which allows us to say that P(x_0, …, x_n | c_i) = P(x_0 | c_i) * P(x_1 | c_i) * … * P(x_n | c_i). This assumption is most likely not true—hence the name *naive* Bayes classifier, but the classifier nonetheless performs well in most situations. Therefore, our final representation of class probability is the following:

$$\begin{aligned} P(c_i|x_0,\ldots,x_n) &\propto P(x_0,\ldots,x_n|c_i)P(c_i) \\ &\propto P(c_i)\prod_{j=1}^{n} P(x_j|c_i) \end{aligned}$$

Calculating the individual P(x_j | c_i) terms will depend on what distribution your features follow. In the context of text classification, where features may be word counts, features may follow a **multinomial distribution**. In other cases, where features are continuous, they may follow a **Gaussian distribution**.



Note that there is very little explicit training in Naive Bayes compared to other common classification methods. The only work that must be done before prediction is finding the parameters for the features' individual probability distributions, which can typically be done quickly and deterministically. This means that Naive Bayes classifiers can perform well even with high-dimensional data points and/or a large number of data points.

2.4 Software

Our project is based on the concept of Data Classification. We require a data set which consists of various factors which cause the success and failure of the movie. Finally we use pandas to access the database and implement our algorithm on the data set to predict.

Required Code Dependencies:

1. Python v3.6
2. Pandas
3. Data Set

2.5 Use- Case Diagram

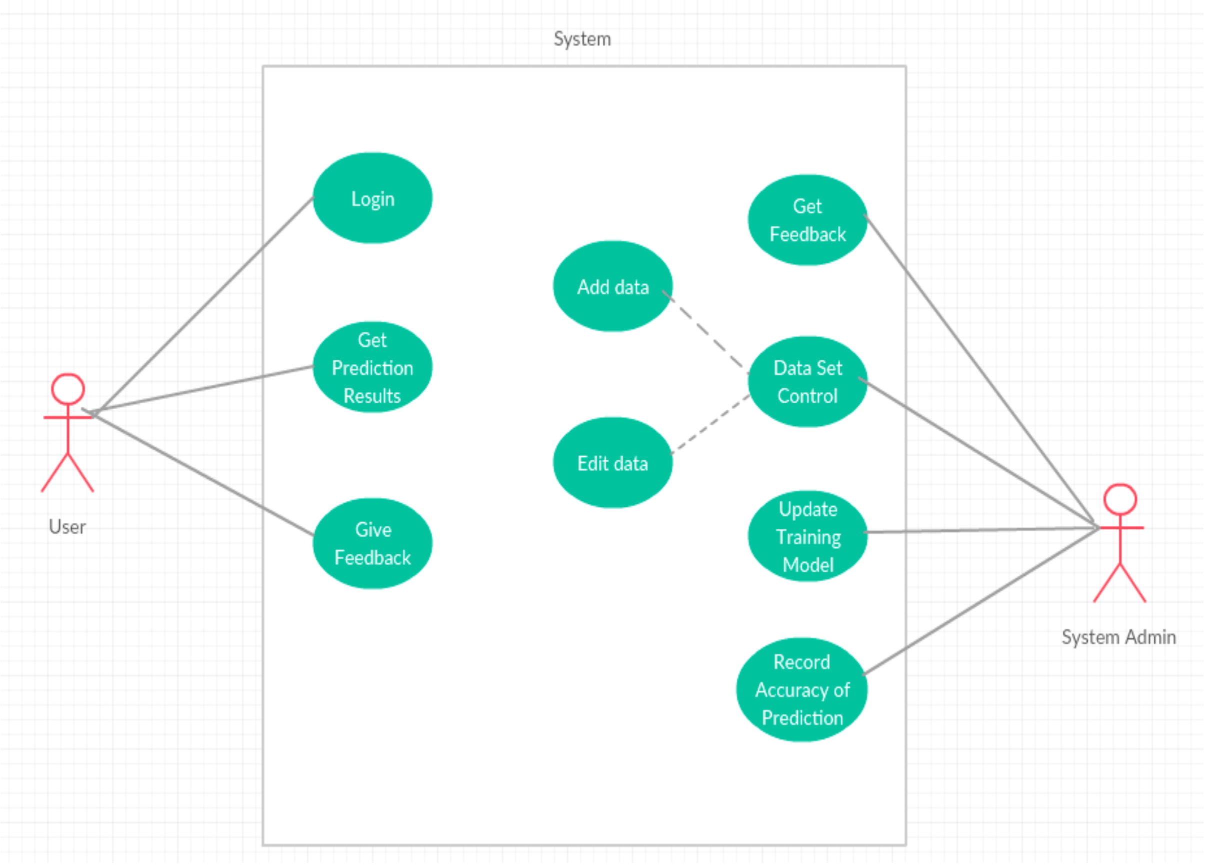


Fig 2.5.1: Use Case Diagram

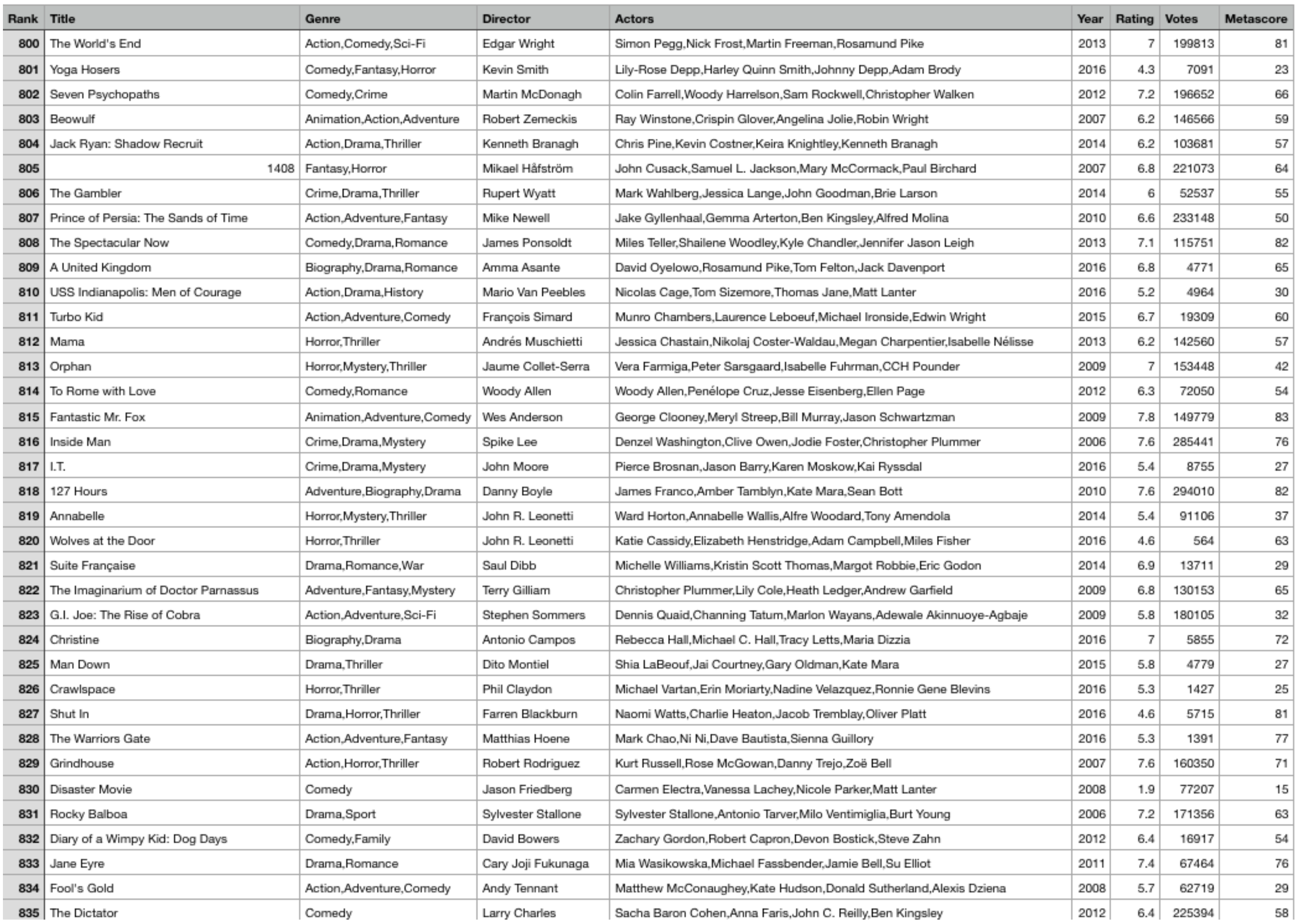
CHAPTER 3

3.1 Implementation

We started off by finding a dataset consisting of entries of movies with various attributes such as Actors, Directors, Genre, etc. The metascore of each such movie entry was found. Movies with missing metascores values were deleted. We then took the metascore and divided into two parts. If the metascore was greater than 65 then it was considered as successful otherwise failure. The data set had multiple entries for its attributes. Hence each input entry needed to be checked with all the entries of a particular attribute. The system works in an interactive way. It first provides the user with an input box and mentions what the user needs to enter. Only after entering this does the system present its next input requirement The user enters a one or multiple values as genre of the movie. These values are taken by the system and probability of each is found. This probability is used as one of the elements in obtaining the final prediction result. Similar operations are performed with the Director and Actor inputs. These inputs provide individual probability and hence work in the prediction of movie success or failure in our Naive Bayes based System. The system has one drawback which is that it cannot predict the success or failure for newcomers. This means that if there is debutant in the movie then the system will not work for it because of the absence of past data for prediction. Hence the system although useful does not work in all scenarios.

We also divided the data set into two one for training and other for testing. We checked the code for accuracy against a testing set to check the credibility and usefulness of the system. The training set consisted of 900 values and the testing set consisted of 100 values. Of these 100 values 32 entries either had a debut director or actor. Hence these could not be considered. The remaining were tested and the results were recorded.

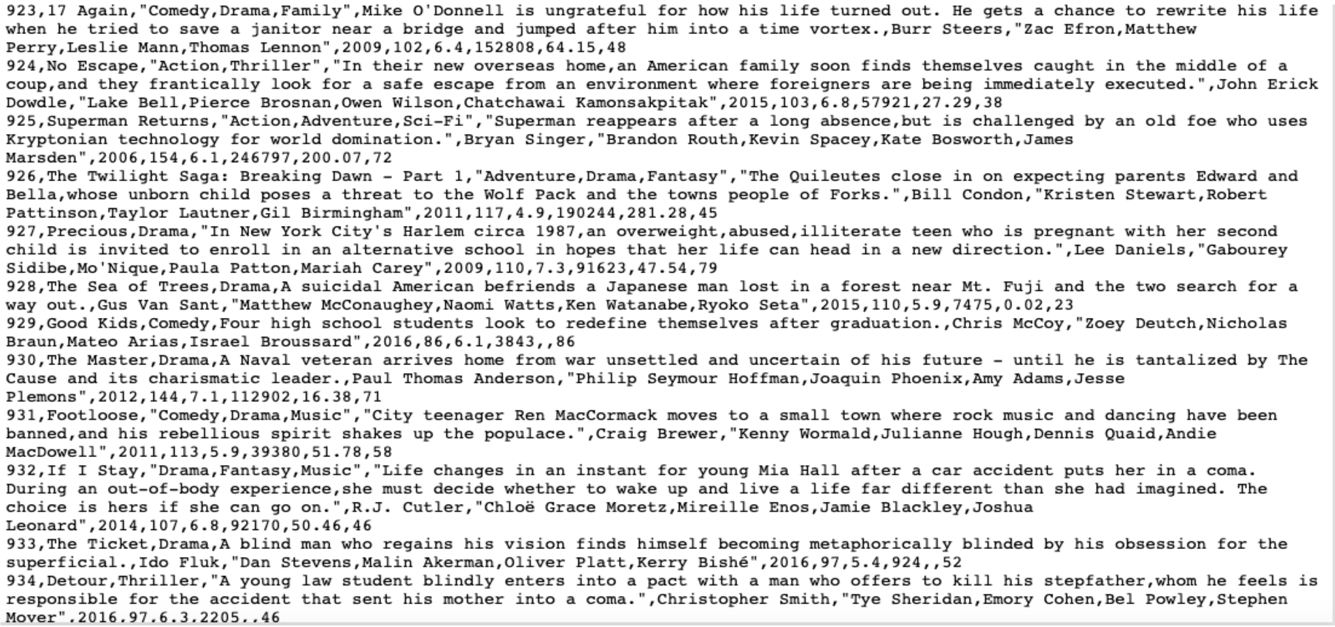
Training Set



Rank	Title	Genre	Director	Actors	Year	Rating	Votes	Metascore
800	The World's End	Action,Comedy,Sci-Fi	Edgar Wright	Simon Pegg,Nick Frost,Martin Freeman,Rosamund Pike	2013	7	199813	81
801	Yoga Hosers	Comedy,Fantasy,Horror	Kevin Smith	Lily-Rose Depp,Harley Quinn Smith,Johnny Depp,Adam Brody	2016	4.3	7091	23
802	Seven Psychopaths	Comedy,Crime	Martin McDonagh	Colin Farrell,Woody Harrelson,Sam Rockwell,Christopher Walken	2012	7.2	196652	66
803	Beowulf	Animation,Action,Adventure	Robert Zemeckis	Ray Winstone,Crispin Glover,Angelina Jolie,Robin Wright	2007	6.2	146566	59
804	Jack Ryan: Shadow Recruit	Action,Drama,Thriller	Kenneth Branagh	Chris Pine,Kevin Costner,Keira Knightley,Kenneth Branagh	2014	6.2	103681	57
805	1408	Fantasy,Horror	Mikael Håfström	John Cusack,Samuel L. Jackson,Mary McCormack,Paul Birchard	2007	6.8	221073	64
806	The Gambler	Crime,Drama,Thriller	Rupert Wyatt	Mark Wahlberg,Jessica Lange,John Goodman,Brie Larson	2014	6	52537	55
807	Prince of Persia: The Sands of Time	Action,Adventure,Fantasy	Mike Newell	Jake Gyllenhaal,Gemma Arterton,Ben Kingsley,Alfred Molina	2010	6.6	233148	50
808	The Spectacular Now	Comedy,Drama,Romance	James Ponsoldt	Miles Teller,Shailene Woodley,Kyle Chandler,Jennifer Jason Leigh	2013	7.1	115751	82
809	A United Kingdom	Biography,Drama,Romance	Amma Asante	David Oyelowo,Rosamund Pike,Tom Felton,Jack Davenport	2016	6.8	4771	65
810	USS Indianapolis: Men of Courage	Action,Drama,History	Mario Van Peebles	Nicolas Cage,Tom Sizemore,Thomas Jane,Matt Lanter	2016	5.2	4964	30
811	Turbo Kid	Action,Adventure,Comedy	François Simard	Munro Chambers,Laurence Leboeuf,Michael Ironside,Edwin Wright	2015	6.7	19309	60
812	Mama	Horror,Thriller	Andrés Muschietti	Jessica Chastain,Nikolaj Coster-Waldau,Megan Charpentier,Isabelle Nélisse	2013	6.2	142560	57
813	Orphan	Horror,Mystery,Thriller	Jaume Collet-Serra	Vera Farmiga,Peter Sarsgaard,Isabelle Fuhrman,CCH Pounder	2009	7	153448	42
814	To Rome with Love	Comedy,Romance	Woody Allen	Woody Allen,Penélope Cruz,Jesse Eisenberg,Ellen Page	2012	6.3	72050	54
815	Fantastic Mr. Fox	Animation,Adventure,Comedy	Wes Anderson	George Clooney,Meryl Streep,Bill Murray,Jason Schwartzman	2009	7.8	149779	83
816	Inside Man	Crime,Drama,Mystery	Spike Lee	Denzel Washington,Clive Owen,Jodie Foster,Christopher Plummer	2006	7.6	285441	76
817	I.T.	Crime,Drama,Mystery	John Moore	Pierce Brosnan,Jason Barry,Karen Moskow,Kai Ryssdal	2016	5.4	8755	27
818	127 Hours	Adventure,Biography,Drama	Danny Boyle	James Franco,Amber Tamblyn,Kate Mara,Sean Bott	2010	7.6	294010	82
819	Annabelle	Horror,Mystery,Thriller	John R. Leonetti	Ward Horton,Annabelle Wallis,Alfre Woodard,Tony Amendola	2014	5.4	91106	37
820	Wolves at the Door	Horror,Thriller	John R. Leonetti	Katie Cassidy,Elizabeth Henstridge,Adam Campbell,Miles Fisher	2016	4.6	564	63
821	Suite Française	Drama,Romance,War	Saul Dibb	Michelle Williams,Kristin Scott Thomas,Margot Robbie,Eric Godon	2014	6.9	13711	29
822	The Imaginarium of Doctor Parnassus	Adventure,Fantasy,Mystery	Terry Gilliam	Christopher Plummer,Lily Cole,Heath Ledger,Andrew Garfield	2009	6.8	130153	65
823	G.I. Joe: The Rise of Cobra	Action,Adventure,Sci-Fi	Stephen Sommers	Dennis Quaid,Channing Tatum,Marlon Wayans,Adewale Akinnuoye-Agbaje	2009	5.8	180105	32
824	Christine	Biography,Drama	Antonio Campos	Rebecca Hall,Michael C. Hall,Tracy Letts,Maria Dizzia	2016	7	5855	72
825	Man Down	Drama,Thriller	Dito Montiel	Shia LaBeouf,Jai Courtney,Gary Oldman,Kate Mara	2015	5.8	4779	27
826	Crawlspace	Horror,Thriller	Phil Claydon	Michael Vartan,Erin Moriarty,Nadine Velazquez,Ronnie Gene Blevins	2016	5.3	1427	25
827	Shut In	Drama,Horror,Thriller	Farren Blackburn	Naomi Watts,Charlie Heaton,Jacob Tremblay,Oliver Platt	2016	4.6	5715	81
828	The Warriors Gate	Action,Adventure,Fantasy	Matthias Hoene	Mark Chao,Ni Ni,Dave Bautista,Sienna Guillory	2016	5.3	1391	77
829	Grindhouse	Action,Horror,Thriller	Robert Rodriguez	Kurt Russell,Rose McGowan,Danny Trejo,Zoë Bell	2007	7.6	160350	71
830	Disaster Movie	Comedy	Jason Friedberg	Carmen Electra,Vanessa Lachey,Nicole Parker,Matt Lanter	2008	1.9	77207	15
831	Rocky Balboa	Drama,Sport	Sylvester Stallone	Sylvester Stallone,Antonio Tarver,Milo Ventimiglia,Burt Young	2006	7.2	171356	63
832	Diary of a Wimpy Kid: Dog Days	Comedy,Family	David Bowers	Zachary Gordon,Robert Capron,Devon Bostick,Steve Zahn	2012	6.4	16917	54
833	Jane Eyre	Drama,Romance	Cary Joji Fukunaga	Mia Wasikowska,Michael Fassbender,Jamie Bell,Su Elliot	2011	7.4	67464	76
834	Fool's Gold	Action,Adventure,Comedy	Andy Tennant	Matthew McConaughey,Kate Hudson,Donald Sutherland,Alexis Dziena	2008	5.7	62719	29
835	The Dictator	Comedy	Larry Charles	Sacha Baron Cohen,Anna Faris,John C. Reilly,Ben Kingsley	2012	6.4	225394	58

Fig 3.1.1: Training Dataset

Testing Set



```
923,17 Again,"Comedy,Drama,Family",Mike O'Donnell is ungrateful for how his life turned out. He gets a chance to rewrite his life
when he tried to save a janitor near a bridge and jumped after him into a time vortex.,Burr Steers,"Zac Efron,Matthew
Perry,Leslie Mann,Thomas Lennon",2009,102,6.4,152808,64.15,48
924,No Escape,"Action,Thriller","In their new overseas home,an American family soon finds themselves caught in the middle of a
coup,and they frantically look for a safe escape from an environment where foreigners are being immediately executed.",John Erick
Dowdle,"Lake Bell,Pierce Brosnan,Owen Wilson,Chatchawai Kamonsakpitak",2015,103,6.8,57921,27.29,38
925,Superman Returns,"Action,Adventure,Sci-Fi","Superman reappears after a long absence,but is challenged by an old foe who uses
Kryptonian technology for world domination.",Bryan Singer,"Brandon Routh,Kevin Spacey,Kate Bosworth,James
Marsden",2006,154,6.1,246797,200.07,72
926,The Twilight Saga: Breaking Dawn - Part 1,"Adventure,Drama,Fantasy","The Quileutes close in on expecting parents Edward and
Bella,whose unborn child poses a threat to the Wolf Pack and the towns people of Forks.",Bill Condon,"Kristen Stewart,Robert
Pattinson,Taylor Lautner,Gil Birmingham",2011,117,4.9,190244,281.28,45
927,Precious,Drama,"In New York City's Harlem circa 1987,an overweight,abused,illiterate teen who is pregnant with her second
child is invited to enroll in an alternative school in hopes that her life can head in a new direction.",Lee Daniels,"Gabourey
Sidibe,Mo'Nique,Paula Patton,Mariah Carey",2009,110,7.3,91623,47.54,79
928,The Sea of Trees,Drama,A suicidal American befriends a Japanese man lost in a forest near Mt. Fuji and the two search for a
way out.,Gus Van Sant,"Matthew McConaughey,Naomi Watts,Ken Watanabe,Ryoko Seta",2015,110,5.9,7475,0.02,23
929,Good Kids,Comedy,Four high school students look to redefine themselves after graduation.,Chris McCoy,"Zoey Deutch,Nicholas
Braun,Mateo Arias,Israel Broussard",2016,86,6.1,3843,,86
930,The Master,Drama,A Naval veteran arrives home from war unsettled and uncertain of his future - until he is tantalized by The
Cause and its charismatic leader.,Paul Thomas Anderson,"Philip Seymour Hoffman,Joaquin Phoenix,Amy Adams,Jesse
Plemons",2012,144,7.1,112902,16.38,71
931,Footloose,"Comedy,Drama,Music","City teenager Ren MacCormack moves to a small town where rock music and dancing have been
banned,and his rebellious spirit shakes up the populace.",Craig Brewer,"Kenny Wormald,Julianne Hough,Dennis Quaid,Andie
MacDowell",2011,113,5.9,39380,51.78,58
932,If I Stay,"Drama,Fantasy,Music","Life changes in an instant for young Mia Hall after a car accident puts her in a coma.
During an out-of-body experience,she must decide whether to wake up and live a life far different than she had imagined. The
choice is hers if she can go on.",R.J. Cutler,"Chloë Grace Moretz,Mireille Enos,Jamie Blackley,Joshua
Leonard",2014,107,6.8,92170,50.46,46
933,The Ticket,Drama,A blind man who regains his vision finds himself becoming metaphorically blinded by his obsession for the
superficial.,Ido Fluk,"Dan Stevens,Malin Akerman,Oliver Platt,Kerry Bishé",2016,97,5.4,924,,52
934,Detour,Thriller,"A young law student blindly enters into a pact with a man who offers to kill his stepfather,whom he feels is
responsible for the accident that sent his mother into a coma.",Christopher Smith,"Tye Sheridan,Emory Cohen,Bel Powley,Stephen
Mover",2016,97,6.3,2205,,46
```

Fig 3.1.2: Testing Dataset

Performance Evaluation:

Classification accuracy alone can be misleading if you have an unequal number of observations in each class or if you have more than two classes in your dataset. A confusion matrix is a technique for summarizing the performance of a classification algorithm. Calculating a confusion matrix can give you a better idea of what your classification model is getting right and what types of errors it is making. A confusion matrix of binary classification is a two by two table formed by counting of the number of the four outcomes of a binary classifier.

Four outcomes – true positive, true negative, false positive and false negative.

- True positive (TP): correct positive prediction
- False positive (FP): incorrect positive prediction
- True negative (TN): correct negative prediction
- False negative (FN): incorrect negative prediction

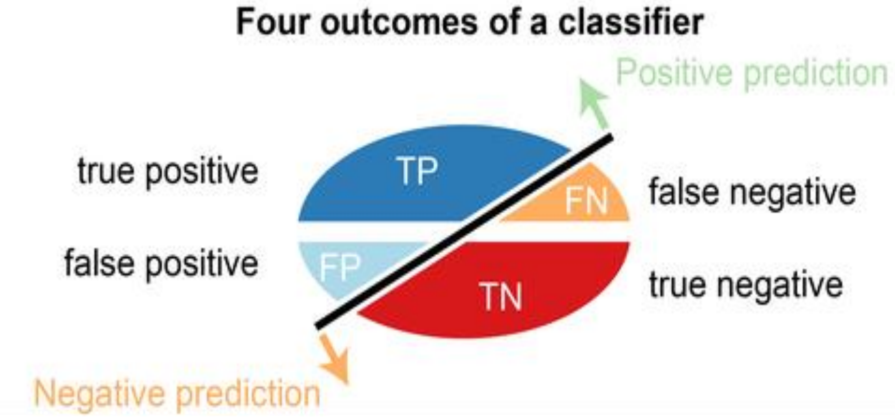


Fig 3.1.3: Outcomes of a classifier

Accuracy is one of the intuitive measures that can be calculated using the confusion matrix.

Accuracy (ACC) is calculated as the number of all correct predictions divided by the total number of the dataset. The best accuracy is 1.0, whereas the worst is 0.0. It can also be calculated by 1 – ERR.

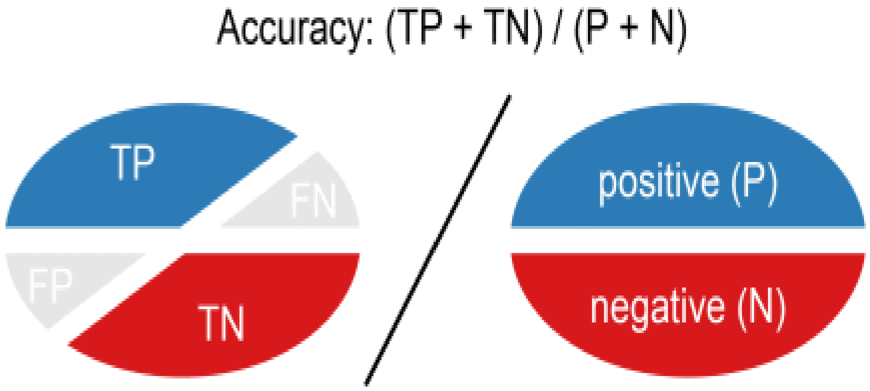


Fig 3.1.4:Accuracy calculation

Accuracy is calculated as the total number of two correct predictions (TP + TN) divided by the total number of a dataset (P + N).

$$ACC = \frac{TP + TN}{TP + TN + FN + FP} = \frac{TP + TN}{P + N}$$

Similarly, based on our project we have created a confusion matrix taking into consideration the testing dataset which consists of 100 entries in order to calculate the efficiency of Naive Bayes Algorithm that we have applied to our dataset.

		Predicted	
		SUCCESSFUL	NOT SUCCESSFUL
Observed	SUCCESSFUL	TP= 24	FN= 0
	NOT SUCCESSFUL	FP= 25	TN= 19

Table 3.1.5: Confusion Matrix

Here,

TP= Movies that were predicted SUCCESSFUL by the system and were actually SUCCESSFUL

TN= Movies that were predicted NOT SUCCESSFUL by the system and were actually NOT SUCCESSFUL

FP= Movies that were predicted SUCCESSFUL by the system but were actually NOT SUCCESSFUL

FN= Movies that were predicted NOT SUCCESSFUL by the system but were actually SUCCESSFUL

Therefore,

$$ACC = \frac{TN + TP}{P + N}$$

$$= \frac{19 + 24}{68}$$

Hence the accuracy provided by the Naive Bayes Classification Algorithm is **0.632**

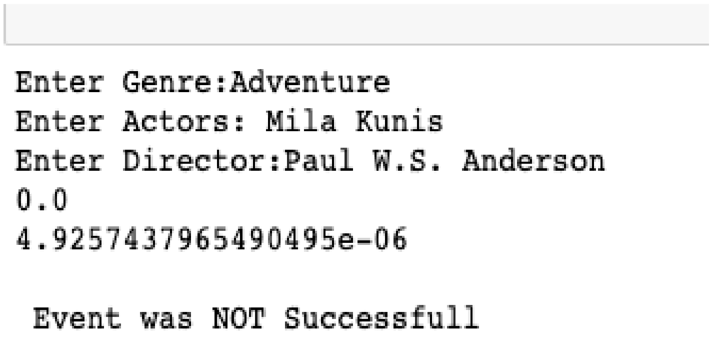
3.2 Algorithm

```
import pandas as pd

data = pd.read_csv("train.csv")
location = data["Genre"].values
event = data["Actors"].values
time = data["Director"].values
success = data["Metascore"].values

l1 = input("Enter Genre:")
l=l1.split(',')
e1 = input("Enter Actors: ")
e=e1.split(',')
d1 = input("Enter Director:")
d=d1.split(',')

numYes = 0
numNo = 0
for i in success:
```

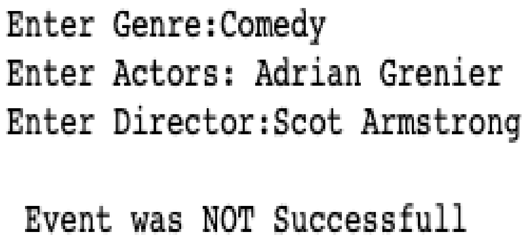
```
if(i >= 65):
numYes = numYes + 1
else:
numNo = numNo + 1
probYes = numYes/data.shape[0]
probNo = numNo/data.shape[0]


numl = 0
nume = 0
numd = 0
for i in location:
    j=i.split(',');
    for k in j:
        for m in l:
            if(k == m):
                numl = numl + 1
                break
probl = numl/data.shape[0]

for i in event:
    j=i.split(',');
    for k in j:
        for m in e:
            if(k == m):
                nume = nume + 1
                break
probe = nume/data.shape[0]

for i in time:
    j=i.split(',');
    for k in j:
        for m in d:
            if(k == m):
                numd = numd + 1
                break
probd = numd/data.shape[0]
numYesl = 0
numNol = 0

for i1,j in zip(location, success):
    i=i1.split(',');
    for k in l:
        for k1 in i:
            if(k1 == k):
                if(j>= 50):
                    numYesl = numYesl + 1
                else:
                    numNol = numNol + 1
probYesl = numYesl/numl
```

```
probNol = numNol/numl


numYese = 0
numNoe = 0
for i1,j in zip(event, success):
    i=i1.split(',')
    for k in e:
        for k1 in i:
            if(k1 == k):
                if(j >= 50):
                    numYese += 1
                else:
                    numNoe += 1
probYese = numYese/nume
probNoe = numNoe/nume


numYesd = 0
numNod = 0
for i1,j in zip(time, success):
    i=i1.split(',')
    for k in d:
        for k1 in i:
            if(k1 == k):
                if(j >= 50):
                    numYesd += 1
                else:
                    numNod += 1
probYesd = numYesd/numd
probNod = numNod/numd

problYes = (probl*probYesl)/probYes
probeYes = (probe*probYese)/probYes
probdYes = (probd*probYesd)/probYes


problNo = (probl*probNol)/probNo
probeNo = (probe*probNoe)/probNo
probdNo = (probd*probNod)/probNo


yes = problYes*probeYes*probdYes
no = problNo*probeNo*probdNo

if(yes >= no):
print("\n Event was Successful")
else:
print("\n Event was NOT Successful")
```

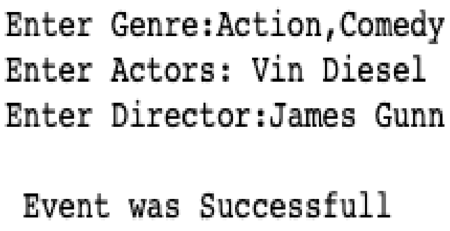
3.3 Screenshots



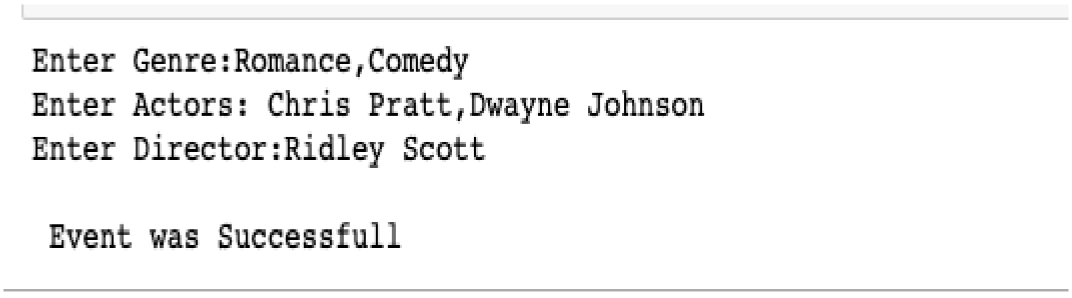
```
Enter Genre:Adventure
Enter Actors: Mila Kunis
Enter Director:Paul W.S. Anderson
0.0
4.925743796549049 5e-06

 Event was NOT Successfull
```



```
Enter Genre:Comedy
Enter Actors: Adrian Grenier
Enter Director:Scot Armstrong

 Event was NOT Successfull
```



```
Enter Genre:Action,Comedy
Enter Actors: Vin Diesel
Enter Director:James Gunn

 Event was Successfull
```



```
Enter Genre:Romance,Comedy
Enter Actors: Chris Pratt,Dwayne Johnson
Enter Director:Ridley Scott

 Event was Successfull
```

CHAPTER 4

4.1 Conclusion

MovePredict is thus a platform which will help the investors who are planning to invest money in movies to prepare themselves enough for the competitive movie world. Investing in a movie is a very difficult task and MoviePredict has been built just to make it much easier for the investment easier.

MoviePredict takes into account past data of various movies and hence determines whether a movies will be successful or not. It takes into consideration various aspects such as Genre, Actors, etc. to determine whether the investment will be successful or not.

The data set will be updated as and when movies are released so that the system becomes better and can be used for various new actors and directors. This will further help in predicting for new movies and result in a wider range values in the data set.

MoviePredict is a system which will improve with time and increase in the size of data set. This system which has an accuracy of 63.2% will results in better choices for the movies and hence result in better cinema.

4.2 Future Scope

Predictions made by MoviePredict can be made more accurate by adding new data as and when new movies released turn out to be successful or failure. Some other points that describe the further scope of this project are:

- Currently, MoviePredict only looks at Hollywood Movies. However, in future we aim at extending the reach for investors who want to get into the regional market i.e. Bollywood, Tollywood, etc. as well.
- Through constant updates the UI of the portal can be improved from time to time to make it more user-friendly and make the user experience better.
- Including Music Videos and other results of events in the entertainment sector will help capturing more of the market and make the system useful for more and more people who want to invest in this sector.

CHAPTER 5

5.1 References

[1] K. R. Apala, M. Jose, S. Motnam, C.-C. Chan, K. J. Liszka, and F. D. Gregorio, “Prediction of movies box office performance using social media,” Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining - ASONAM '13, 2013.

[2] S. Gopinath, P. K. Chintagunta, and S. Venkataraman, “Blogs, Advertising, and Local-Market Movie Box Office Performance,” Management Science, vol. 59, no. 12, pp. 2635–2654, 2013.

[3] M. C. A. Mestyán, T. Yasseri, and J. Kertész, “Early Prediction of Movie Box Office Success Based on Wikipedia Activity Big Data,” PLoS ONE, vol. 8, no. 8, 2013.

[4] M.H Latif, H. Afzal “Prediction of Movies popularity Using Machine Learning Techniques”, National University of Sceinces and technology, H-12, ISB, Pakistan.

[9] P. Chaovalit and L. Zhou, “Movie review mining: a comparison between supervised and unsupervised classification approaches,” in Proceedings of the Hawaii International Conference on System Sciences (HICSS), 2005.

CHAPTER 6

6.1 Acknowledgement

Our team is grateful for being given an opportunity to build a project in the domain of Machine Learning and would like to thank all the people who have given us their co-operation in making this project a success. We extend our sincere thanks to Mrs. Shilpa Verma, our lab teacher and project co-ordinator for providing us with her constant guidance and encouragement. This helped us complete our project in machine learning. We also would like to show our appreciation for our college Thadomal Shahani College of Engineering and our entire faculty for their constant guidance and support which has helped us reach where we are and enabling us to use our theoretical knowledge in practical applications in various domains.