

Prediction of Movie Success Using Classification

* *Ritu Jhalani*

** *Harshita Virwani*

*** *Divyanshi Goyal*

**** *Somya Vashishtha*

Abstract

In the film industry, the largest producer of films in the world is India. The Indian film industry was established in 1913 and is the second oldest in the world. India was the third largest with box office revenue of US \$ 2.18 billion in 2017. The Indian film industry is multi-lingual. Hindi film Industry is the largest film industry in India and is mostly based in Mumbai (Bombay), which is referred to as "Bollywood". This paper attempts to predict whether an upcoming movie would be a blockbuster, neutral or a flop. By predicting this, it can help production houses in advertising and to find the best time period to release a movie by looking at the overall environment. This paper proposes making use of classification technique of Data Mining i.e. Naïve Bayes Theorem and Decision Tree on Data Mining Tool named Orange. Data mining is a process to transform raw data into useful information. By applying data mining, we can discover a large set of patterned data. Machine Learning Statistics and Database Systems are involved in Data Mining. In Knowledge Discovery (KDD) process, data mining is an analysis step. Classification helps us to classify the data according to the attributes of the data with respect to a predefined set of classes. Naïve Bayes is a theorem of data mining which is able to predict categorical class labels (blockbuster, neutral, and flop) that classifies data according to rating, month, year of release, genres such as drama, action, romance, comedy, mystery, thriller, and other attributes, and values to classify an upcoming movie. Decision tree helps in supervised learning by creating a training model which can help us predict class values by learning decisions from prior data. To perform the research, we used a Data Mining Tool named Orange which is an open source component-based Visual Programming Software Package for data visualization, machine learning, data mining, and data analysis.

Keywords: Classification, data mining, decision tree, Naïve Bayes, Orange tool

I. INTRODUCTION

As India is the largest producer of films in the world, there is an abundance of movies being released each year. This creates the need for production houses to know if they should make a movie, the chances of the movie being a blockbuster, hit, neutral or flop, or the best time of release for a movie, as well as for the patrons to know which movie to watch or not or when will the best movies will be releasing.

However, there is no pre-existing data repository for Indian cinema, so for production houses and patrons, it can be difficult to predict such. If there is a data

repository, we can use it to understand the data and derive valuable information from it before making or producing a movie as it will be a crucial step for the entertainment industry to ensure the success of a movie that it is investing in. The success of movies is predicted on the basis of multiple criteria. Some criteria are the director, producer, actress, actor, the release date, music director, writer etc. To predict the success of a movie, we used Naïve Bayes method which comes under the classification technique. We implemented Naïve Bayes method on the data mining tool Orange. Data mining can help us extract the information needed to determine if a movie will be successful or not. Classification method

Manuscript received September 2, 2018; revised September 20, 2018; accepted October 10, 2018. Date of publication November 6, 2018.

* R. Jhalani is Assistant Professor (Computer Science) with International School Of Informatics & Management, Sector-12, Mahaveer Marg, Mansarovar, Jaipur, Rajasthan, India-302020. (email : ritu.jhalani87@gmail.com)

** H. Virwani is M.C.A. student at The IIS University, Sector-12, Mahaveer Marg, Mansarovar, Jaipur, Rajasthan, India-302020. (email: harshitavirwani@gmail.com)

*** D. Goyal is M.C.A. student at The IIS University, Sector-12, Mahaveer Marg, Mansarovar, Jaipur, Rajasthan- 302020. (email: Italicakyuso@gmail.com)

**** S. Vashishtha is M.C.A. student at The IIS University, Sector-12, Mahaveer Marg, Mansarovar, Jaipur, Rajasthan, India -302020. (email: somyavashishtha@gmail.com)

DOI: 10.17010/ijcs/2018/v3/i6/141443

Naïve Bayes can then be implemented on the data using Orange to predict the success of the movie.

II. LITERATURE REVIEW

To get prediction names of Bollywood movies were collected [1, 2]. Many researchers have carried out prediction of movies by using different approaches, for example, predicting movie success based on IMDB Data by Nitin Vr. Sarah Babu Pb in the year 2014 by using regression techniques [3]. They applied different models and found out that the linear model represents the movie feature more accurately. The Support Vector Machines (SVM) technique was used by Jeffrey Ericson and Jesse Grodman for predicting movie success which revealed that movies with the words "life" or "story" in the description have higher ratings [4]. Prediction of movie success for real-world movie dataset by Sanjay Pramod, Abhisht Joshi, and Geeta Mary used Fuzzy Logic which has more than 85% accuracy under true positives with Fuzzy Logic [5]. Simple logistic and logistic regression techniques were used by Latif and Afzal for the prediction of movies and they found that the attributes that contributed the most to information are metal core and number of votes for each movie [6]. Saree, White, and Eccleston realized that the Universal Classifier Query proved to be more accurate which is capable of making suggestions for a movie in pre-production [7].

III. DATA COLLECTION, INTEGRATION, AND TRANSFORMATION

The initial step was data collection. There was no pre-existing dataset available for Bollywood movies, so we merged data from multiple sources. Our main sources of data were Wikipedia [1, 2] and box office analysis. The data were collected and merged according to directors and the movies released in previous 10 years. We obtained attributes such as name, genre, director, producer, writer, actor, actress, production houses, and music director of a movie from Wikipedia, and from box office collection the attributes that were in consideration were budget, and movie-success.

Data obtained from different resources, i.e. Wikipedia, box office analysis, etc. were then integrated into one database. Fig.1 shows steps in which data is merged are remodeled so that the classification process would be more productive and adequate.

Here, movie-success is the target variable, so the

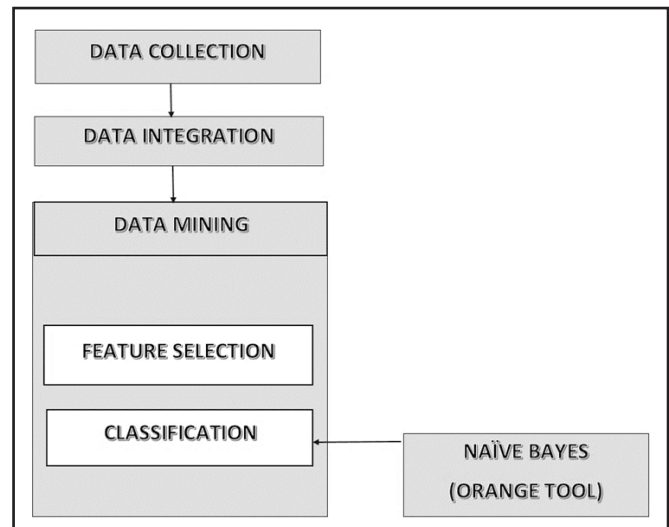


Fig. 1. Data Processing Steps

movies can be recognized as blockbuster, hit, average, or flop.

IV. TECHNOLOGIES

A. Data Mining

Data mining is a process that sorts a large set of data to identify patterns and establish relations to solve problems through data analysis. Data mining tools such as Orange allow enterprises to predict future trends. Data mining techniques are used in many research areas such as mathematics, cybernetics, genetics, and marketing. While data mining techniques are a means to drive efficiencies and predict customer behavior; if used correctly, a business can set itself apart from its competition through the use of predictive analysis. Data mining benefits come from the ability to uncover hidden patterns and relationships in data that can be used to make a prediction that impacts businesses. Specific benefits differ depending on the goal and the industry. Sales and marketing departments can mine customer data to improve lead conversion rates or to create one-to-one marketing campaigns. Data mining information on historical sales patterns and customer behavior can be used to build prediction models for future sales, new products, and services. Companies in the financial industry use data mining tools to build risk models and detect fraud. The manufacturing industry uses it to improve product safety, identify quality issues, manage the supply chain, and improve operations.

B. Classification

Classification is a data mining technique that assigns

data items in a collection to target categories or classes. It accurately predicts the target class for each case in the data. Classification is intended for the analysis of very large datasets to be effective. A dataset with class assignments is needed first for the classification. Discretion is provided with classification as it does not imply order. The classification has many implications for customer segmentation, business modeling, marketing, credit card analysis, and biomedical, and drug response modeling.

C. Naïve Bayes

Naïve Bayes is a classification method for binary or multi-class classification problems. The technique is the easiest to understand when described using categorical input values. Naïve Bayes provides us with a way with which we can calculate the probability of a given hypothesis. It does not consider the dependencies between attributes. The calculation of the probabilities for each hypothesis is simplified to make their calculation traceable.

D. Orange

Orange is a data mining tool which is open source data visualization, machine learning, and data mining toolkit. Orange has a visual programming front-end for explorative data analysis and interactive data visualization. Orange uses Python open-source libraries for scientific computing and its GUI operates within the cross-platform QT framework.

V. EXPERIMENT USING ORANGE TOOL AND RESULTS

Orange is an open source data mining tool which is useful for interactive data data visualization and explorative data analysis of simple and complex problems with additions or widgets.

New data mining workflows can be created with the Orange tool. Data mining workflows comprise of computational components called widgets. The work and exchange of information are fulfilled by widgets. In

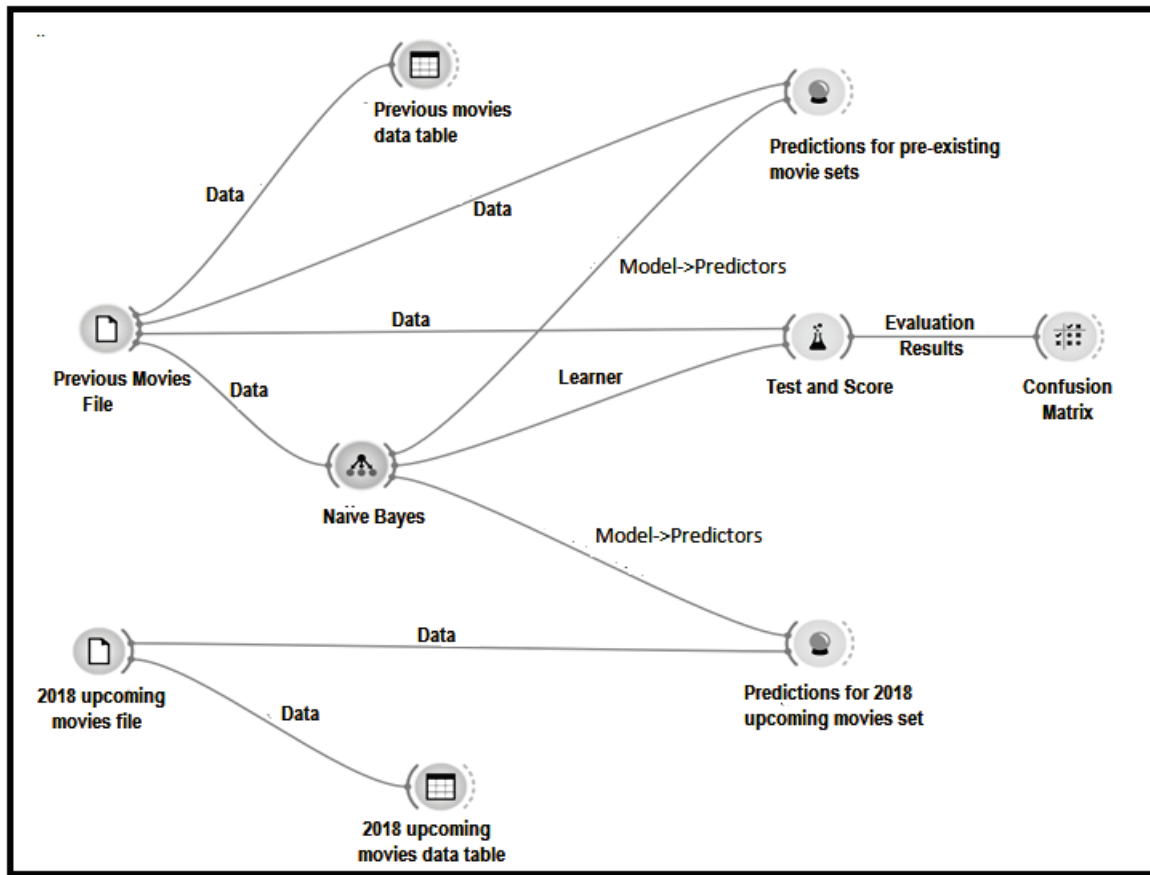


Fig. 2. Workflow Created for the Prediction of Success of a Movie

the workflow shown in fig. 2, the *previous movies* file contains data that have been collected and integrated from various sources and it contains data of released movies according to different attributes such as:

TABLE I. TYPE WITH ATTRIBUTES

TYPE	ATTRIBUTES
Categorical	Genre, Movie-Success, Director, Writer, Actor, Actress, Music-Director
Numerical	Budget
Text	Name, Producer, Production-house

In Table I, *type with attributes* consists of categorical, numerical, and text types. Categorical types are the types in which attributes have different categorical values. For example, for genre we have different categorical values such as action, anthology, black comedy, Bollywood, comedy, crime, dance, drama, fantasy, horror, mystery, political, romance, sports, and thriller. Numeric types are those which have numeric values such as budget of the movie. The text type has some additional information about the entity and is not in categorical form as it has its own individual existence.

In Table II, role with attributes consists of Meta, feature, and target attributes. Meta is a descriptive field for name, producer, and writer of the movie, and production-house which is not used in modeling but will be an aid for identification or as additional information.

TABLE II. ROLE WITH ATTRIBUTES

ROLE	ATTRIBUTES
Feature	Genre, Budget Director, Actor, Actress, Music-Director
Meta	Name, Producer, Production-house, Writer
Target	Movie-Success

The feature is dependent on fields from which the target value is decided, for example, success of a movie is determined by genre, budget, director, actor, actress, and music-director of the movie. Target values are decided by feature value, for example, whether a movie is a blockbuster, hit, neutral, or flop is determined by feature values such as genre, budget, director, actor, actress, and music-director of the movie. In previous movie data table, we can view the contents of the previous movie file.

Naïve Bayes is a fast and simple probabilistic predictor based on Bayes Theorem. Previous movie file is taken as input for Naïve Bayes widget. To view the predictions, the Naïve Bayes serves as input for predictions for the previous movie set and different predictions as shown in fig.2.

In fig. 3 we can see that for Bol Bachchan movie there are 27% chances for blockbuster, 5% chances are there for a flop, 65% chances are there for hit, and 4% chances are there for neutral. So, the highest chances are of 'hit' and the result according to Naïve Bayes is a 'hit'.

In 2018, upcoming movies file contains the file for

	Naïve Bayes	MOVIE-SUCCESS	NAME	PRODUCER	WRITER	DUCTION - HOU	GENRE	GET(in cr)	DIRECTOR	ACTOR	ACTRESS	MUSIC DIRECTOR
10	0.16:0.04:0.77:0.03 - Hit	Hit	Don 2	Farhan Akhtar	Farhan Akhtar	Excel Entertainm...	Crime	85.00	Farhan Akhtar	Shah Rukh Khan	Priyanka Chopra	Shankar-Ehsaan...
11	0.16:0.04:0.77:0.03 - Hit	Hit	Don	Ritesh Sidhwani	Farhan Akhtar	Excel Entertainm...	Crime	380.00	Farhan Akhtar	Shah Rukh Khan	Priyanka Chopra	Shankar-Ehsaan...
12	0.10:0.03:0.15:0.72 - Neutral	Neutral	Lakshya	Ritesh Sidhwani	Farhan Akhtar	Excel Entertainm...	Drama	32.00	Farhan Akhtar	Hritvik Roshan	Priety Zinta	Shankar-Ehsaan...
13	0.44:0.39:0.16:0.01 - BlockBuster	Hit	Ae Dil Hai Mus...	Apoorva Mehta	Karan Johar	Dharma Product...	Romantic	80.00	Karan Johar	Ranbir Kapoor	Anushka Sharma	Pritam
14	0.52:0.20:0.25:0.03 - BlockBuster	Hit	Kabhi Alvida N...	Hiroo Yash Johar	Karan Johar	Dharma Product...	Romantic	50.00	Karan Johar	Shah Rukh Khan	Rani Mukerji	Shankar-Ehsaan...
15	0.99:0.01:0.00:0.00 - BlockBuster	BlockBuster	Kabhi Khushi K...	Yash Johar	Karan Johar	Dharma Product...	Drama	45.00	Karan Johar	Shah Rukh Khan	Kajol	Jaatin Lalit
16	0.98:0.01:0.01:0.00 - BlockBuster	BlockBuster	Kuch Kuch Hot...	Yash Johar	Karan Johar	Dharma Product...	Romantic	10.00	Karan Johar	Shah Rukh Khan	Kajol	Jaatin Lalit
17	0.97:0.01:0.01:0.01 - BlockBuster	BlockBuster	my name is kha...	Hiroo Yash Johar	Karan Johar	Dharma Product...	Drama	45.00	Karan Johar	shah Rukh Khan	Kajol	Shankar-Ehsaan...
18	0.33:0.18:0.27:0.21 - BlockBuster	Hit	Student of the ...	Hiroo Yash Johar	Niranjay Iyengar	Dharma Product...	Drama	59.00	Karan Johar	Sidharth Malhot...	Alia Bhatt	Vishal-Shekhar
19	0.44:0.08:0.23:0.25 - BlockBuster	Neutral	All the Best: Fu...	Ajay Devgn	Rohit Shetty	Ajay Devgn Films	Comedy	40.00	Rohit Shetty	Ajay Devgn	Bipasha Basu	Pritam
20	0.27:0.05:0.65:0.04 - Hit	Hit	Bol Bachchan	Ajay Devgn	Rohit Shetty	Ajay Devgn Films	Comedy	66.00	Rohit Shetty	Ajay Devgn	Asin	Himesh Resham...
21	0.08:0.01:0.90:0.01 - Hit	Hit	Chennai Express	Gauri Khan	K. Subash	Red Chillies Ente...	Comedy	115.00	Rohit Shetty	Shah Rukh Khan	Deepika Paduk...	Vishal-Shekhar
22	0.29:0.12:0.58:0.01 - Hit	Hit	Dilwale (2015)...	Gauri Khan	Gauri Khan	Red Chillies Ente...	Romantic	165.00	Rohit Shetty	Shah Rukh Khan	kajol	Pritam
23	0.40:0.02:0.50:0.08 - Hit	BlockBuster	Gomaal 3	Dhillin Mehta	Sajid-Farhad	Eros International	Comedy	30.00	Rohit Shetty	Ajay Devgn	Kareena Kapoor	Pritam
24	0.81:0.04:0.12:0.03 - BlockBuster	BlockBuster	Gomaal Again	Rohit Shetty	Sajid-Farhad	Rohit Shetty Pro...	Comedy	80.00	Rohit Shetty	Ajay Devgn	Parineeti Chopra	Amar Mohile
25	0.40:0.02:0.50:0.08 - Hit	Hit	Gomaal Returns	Dhillin Mehta	Sajid-Farhad	Eros International	Comedy	35.00	Rohit Shetty	Ajay Devgn	Kareena Kapoor	Pritam
26	0.32:0.03:0.13:0.51 - Hit	Hit	Gomaal: Fun U...	Dhillin Mehta	Neeraj Vora	Shree Aashwinay...	Comedy	15.00	Rohit Shetty	Ajay Devgn	Rimi Sen	Vishal-Shekhar
27	0.23:0.06:0.63:0.08 - Hit	Hit	Singham	Maheesh Rama...	Farhan Akhtar	Reliance Entertai...	Action	55.00	Rohit Shetty	Ajay Devgn	Kajal Aggarwal	Ajay-Atul
28	0.28:0.02:0.68:0.02 - Hit	Hit	Singham Returns	Ajay Devgn	Sajid-Farhad	Ajay Devgn Films	Action	95.00	Rohit Shetty	Ajay Devgn	Kareena Kapoor	Ankit Tiwari
29	0.93:0.01:0.01:0.05 - BlockBuster	BlockBuster	Taare Zameen ...	Aamir Khan	Amole Gupte	Aamir Khan Prod...	Drama	12.00	Aamir Khan	Aamir Khan	Tisca Chopra	Shankar-Ehsaan...
30	0.01:0.07:0.87:0.05 - Hit	Hit	Bajirao Mastani	Sanjay Leela B...	Sanjay Leela B...	Bhansali Product...	Romantic	145.00	Sanjay Leela Bh...	Ranveer Singh	Deepika Paduk...	Sanjay Leela Bh...
31	0.02:0.09:0.02:0.87 - Neutral	Neutral	Black	Sanjay Leela B...	Sanjay Leela B...	SLB Films	Drama	18.00	Sanjay Leela Bh...	Amitabh Bachc...	Rani Mukerji	Monty Sharma

Fig. 3. Prediction for Previous Movies Set by Naïve Bayes

which the results are to be predicted and that contains the datasets for 2018 upcoming movies. We have same values for different attributes such as previous movies file.

In 2018 upcoming movies data table we can view the contents of the 2018 upcoming movies file. 2018 upcoming movie file is attached to the predictions for 2018 upcoming movies set and the result for this file is predicted by Naïve Bayes method according to previous movies file and the results for the predictions are discussed here.

In fig. 4, for Pari movie, we can see that there are 14% chances for blockbuster, 17% chances are there for flop, 18% chances are there for hit, and 52% chances are there for neutral. So, the highest chances are of neutral and the result according to Naïve Bayes is neutral according to the previous movie dataset.

In fig.5, Confusion matrix shows the number/proportion of instances between the predicted and actual class. In this way, one can observe which specific instances are misclassified and how. In the figure, purple background cells are correctly classified and others are misclassified. For correctly classified, 18 instances were blockbuster and prediction of these are also blockbuster and for misclassified, 7 instances were there for hit but were predicted as Blockbuster.

VI. CONCLUSION

The proposed research aims to predict the success of movies on different factors such as genre, budget, director, actor, actress, and production-house. We have used Naïve Bayes for prediction for our experimentation. Two predictions were made. One, we predicted the

Info	Naive Bayes	NAME	PRODUCER	WRITER	DOCUCTION - HOL	GENRE	BUDGET(in crores)	DIRECTOR	ACTOR	ACTRESS	MUSIC DIRECTOR
Data: 49 instances. Predictors: 1 Task: Classification	1 0.26 : 0.16 : 0.13 : 0.44 -- Neutral	1921.0	Vikram bhatt	Vikram bhatt	Motion Picture ...	Horror	15.00	Vikram bhatt	Karan Kundra	Zareen Khan	Harish Sagane
<input type="button" value="Restore Original Order"/>	2 0.21 : 0.12 : 0.34 : 0.33 -- Hit	Kaalakaandi	Rohit Khattar	Akshat Verma	Cinestaan Film ...	Drama	25.00	Akshat Verma	Saif Ali Khan	Sobhita Dhulip...	Sameer Uddin
Show	3 0.39 : 0.11 : 0.23 : 0.27 -- BlockBuster	Mukkabaz	Anand L Rai	Anurag Kashyap	Phantom Films	Drama	10.00	Anurag Kashyap	Vineet Kumar Si...	Zoya Hussain	Prashant Pillai
<input checked="" type="checkbox"/> Predicted class	4 0.23 : 0.15 : 0.23 : 0.39 -- Neutral	My Birthday Song	Sanjay Suri	Samir Soni	Kahwa Entertai...	Thriller	3.00	Samir Soni	Sanjay Suri	Suparna Krishna	Rajeev V Bhalla
<input checked="" type="checkbox"/> Predicted probabilities for:	5 0.23 : 0.15 : 0.23 : 0.39 -- Neutral	Vodka Diaries	Vishal Karkera	Vaibhav Bajpai	K'Scope Enterta...	Thriller	1.26	Kushal Srivastava	Vineet Kumar Si...	Raima Sen	Sandesh Shandi...
<input type="button" value="BlockBuster"/>	6 0.23 : 0.15 : 0.23 : 0.39 -- Neutral	Nirdosh	Pradeep Rangw...	Amit Khan	UV Films	Thriller	0.60	Pradeep Rangw...	Arbaaz Khan	Maheek Chahal	Uyaqat Ajmeri
<input type="button" value="Flop"/>	7 0.39 : 0.11 : 0.23 : 0.27 -- BlockBuster	Union Leader	Sanjay Patel	Sanjay Patel	Dimlight Pictures	Drama	0.08	Sanjay Patel	Rahul Bhat	Tillotama Shome	ankit tiwari
<input type="button" value="Hit"/>	8 0.26 : 0.14 : 0.47 : 0.12 -- Hit	Padmaavat	Sanjay Leela Bh...	Sanjay Leela Bh...	Bhansali Produc...	Drama	215.00	Sanjay Leela Bh...	Shahid Kapoor	Deepika Paduk...	Sanjay Leela Bh...
<input type="button" value="Neutral"/>	9 0.34 : 0.21 : 0.17 : 0.28 -- BlockBuster	Wo India Ka Sh...	Jagdish Vishwak...	Suresh Mandal	Impress Media	Suspense	10.00	Suresh Mandal	Raza Murad	Yogita Rajput	Sachin Pathak
<input checked="" type="checkbox"/> Draw distribution bars	10 0.31 : 0.17 : 0.27 : 0.24 -- BlockBuster	Pad Man	Twinkle Khanna	R.Balki	Columbia Pictur...	Drama	55.00	R. Balki	Akshay Kumar	Sonam Kapoor	Amit Trivedi
Data View	11 0.36 : 0.15 : 0.30 : 0.20 -- BlockBuster	Love Per Square...	Ronnie Screwala	Anand Tiwari	RSVP Production	Comedy	5.00	Anand Tiwari	Vicky Kaushal	Supriya Pathak	Sohail Sen
<input checked="" type="checkbox"/> Show full dataset	12 0.18 : 0.21 : 0.27 : 0.34 -- Neutral	Aiyaary	Shital Bhatia	Neeraj Pandey	Motion Picture ...	Thriller	65.00	Neeraj Pandey	sidharth malhotra	Rakul Preet Singh	Sanjay Chowdh...
Output	13 0.25 : 0.32 : 0.30 : 0.13 -- Flop	Kuch Bheeg...	Vikram Mehra	Abhishek Chatter...	Dimlight Pictures	Romantic	10.00	Onir	Shray Tiwari	Geetanjali Thapa	Shashwat Srivas...
<input checked="" type="checkbox"/> Original data	14 0.19 : 0.16 : 0.42 : 0.24 -- Hit	Sonu Ke Titu Ki ...	Bhushan Kumar	Rahul Mody	T-Series	Comedy	30.00	Luv Ranjan	Kartik Aaryan	Nushrat Bharucha	Rochak Kohli
<input checked="" type="checkbox"/> Predictions	15 0.19 : 0.16 : 0.42 : 0.24 -- Hit	Welcome To Ne...	Vashu Bhagnani	Dheeraj Rattan	Wiz Films	Comedy	26.00	Chakri Toleti	Diljit Dosanjh	Sonakshi Sinha	Sajid-Wajid
<input checked="" type="checkbox"/> Probabilities	16 0.14 : 0.17 : 0.18 : 0.52 -- Neutral	pari	Anushka Sharma	Prosit Roy	Clean Slate Films	Horror	18.00	Prosit Roy	Parambrata Cha...	Anushka Sharma	Anupam Roy
	17 0.19 : 0.16 : 0.42 : 0.24 -- Hit	Veerey Ki Wedd...	Rajat Bakshi	Amit Khan	Make My Day E...	Comedy	20.00	Ashu Trikha	Pulkit Samrat	Kriti Kharbanda	Meet Bros
	18 0.11 : 0.14 : 0.31 : 0.44 -- Neutral	Hate Story 4	Bhushan Kumar	Milap Milan Za...	T-Series Films	Thriller	22.04	Vishal Pandya	Vivan Bhatena	Urvashi Rautela	Sunny and Inder...
	19 0.19 : 0.16 : 0.42 : 0.24 -- Hit	Dil Jaunglee	Deepshikha Des...	Aleya Sen	Pooja Entertain...	Comedy	20.00	Aleya Sen	Saqib Saleem	Taapsee Pannu	Tanishk Bagchi
	20 0.39 : 0.11 : 0.23 : 0.27 -- BlockBuster	3 Storeys	Ritesh Sidhwani	Arjun Mukerjee	Excel Entertain...	Drama	10.00	Arjun Mukerjee	Sharmar Joshi	Richa Chadda	Clinton Cerejo
	21 0.18 : 0.21 : 0.27 : 0.34 -- Neutral	raid	Abhishek Pathak	Ritesh Shah	Panorama Studi...	Thriller	45.00	Raj Kumar Gupta	Ajay Devgn	Ileana D'Cruz	Amit Trivedi
	22 0.21 : 0.12 : 0.34 : 0.33 -- Hit	nicki	Maneesh Sharma	Ankur Chaudhry	Yash Raj Films	Drama	20.00	Siddharth P. Ma...	Shivkumar Subr...	Rani Mukerji	Jasleen Royal

Fig. 4. Prediction for 2018 Upcoming Movies Set by Naïve Bayes

		Predicted				
		BlockBuster	Flop	Hit	Neutral	Σ
Actual	BlockBuster	18	0	1	0	19
	Flop	0	19	0	0	19
	Hit	7	13	18	6	44
	Neutral	3	2	0	15	20
Σ		28	34	19	21	102

Fig. 5. Confusion Matrix

previous year dataset to see how accurate results were according to different parameters, and the second prediction was made about the status of the movie, whether it would be a blockbuster, a hit, a neutral, or a flop according to previous year data set. After classification, we found Naive Bayes provided 97% accuracy on the success of movie being a blockbuster, hit, neutral or flop.

REFERENCES

- [1] "List of Bollywood films of 2018," Wikipedia. [Online]. Available: https://en.wikipedia.org/wiki/List_of_Bollywood_films_of_2018
- [2] "List of Bollywood films," Wikipedia. [Online]. Available: https://en.wikipedia.org/wiki/Lists_of_Bollywood_films
- [3] V. R. Nithin, M. Pranav, P. B. Sarath, and Lijiya, "Predicting movie success based on IMDB Data," 2014.
- [4] J. Ericson and J. Goodman, "A predictor for movie success," 2013.
- [5] S. Pramod, A. Joshi and A. G. Mary, "Prediction of movie success for real world movie dataset," *Int. J. of Advance Res., Ideas and Innovations in Technol.*, vol. 3, no. 3, 2017. [Online]. Available: <https://www.ijariit.com/manuscripts/v3i3/V3I3-1228.pdf>
- [6] M. H. Latif and H. Afzal, "Prediction of movies popularity using machine learning techniques," *Int. J. of Comput. Sci. and Network Security*, vol. 16, no. 8, 127-131, 2016.
- [7] M. Saraee, S. White and J. Eccleston, "A data mining approach to analysis and prediction of movie ratings," in *The 5th Int. Conf. on Data Mining, Text Mining, and their Bus. Appl.*, 2004. [Online]: Available: http://usir.salford.ac.uk/18838/1/Wessex_movie.pdf

About the Authors



Ritu Jhalani is Assistant Professor (Computer Science) with International School Of Informatics & Management. She completed Masters of Computer Applications (M.C.A.) from BIT, Mesra. She has five and a half years of work experience. Her research areas are data mining, web technology, and mining tool.



Harshita Virwani is an M.C.A student at the IIS University.



Divyanshi Goyal is an MCA Student at the IIS University.



Somya Vashishtha is an MCA Student at The IIS University.