OPINION MINING APPROACH TO EVALUATE AND PREDICT THE TELEVISION RATING POINT OF SERIALS

Karamjeet Kaur[*]

Seema Baghla[*]

Abstract:

Opinion mining has been used for tracking the mood of people about a particular product or service. It is being used as an important tool for judgment of most viewed programs on various channels. In present work, opinion mining has been used to find television rating point (TRP) of various serials on national channels. In this work, the dataset has been prepared by extracting the information from the historical records. Random tree classification technique has been applied for prediction of TRP of TV serials. Six major factors gender, age, qualification, marital status, occupation and area have been analyzed for prediction of TRP. WEKA tool has been used for classification and classification of the gathered data. The TRP of the channels has been predicted in terms of responses such as analysis, accuracy and predictor. The TRP of the serials has been calculated on the basis of opinion mining. On the basis of the results of the present work, it has been concluded that the TRP of the rating of serials is positive or negative

Key words: Data Mining, Opinion Mining, TRP, Classification, Prediction.

[*] M. Tech. Student (Comp. Engg), Assistant Professor (Comp. Engg.) Supervisor,Yadavindra College of Engineering, Punjabi University Guru Kashi Campus, Talwandi Sabo, Bathinda, Punjab, India

I. INTRODUCTION

Data mining is described as extracting the useful information from huge set of data. In other words data mining is the procedure of mining knowledge from data. Data mining also involves other processes such as data cleaning, data integration, data transformation, pattern evaluation and data representation. It is one of the tasks in the process of knowledge discovery from the database. Two goals of data mining are classification and prediction. Prediction tells us about the unknown value of future variables.

TRP is Television Rating Point. It is the tool that tells us which channel and the program is viewed most or it indicates the popularity

of a TV channel or a program. TRP shows how many times people are watching a channel or a particular program. TRP enables the advertisers and investors to understand the mood of the people. TRP is calculated by Indian agencies namely INTAM and DART. INTAM is Indian Television Audience Measurement. DART is Doordarshan Audience Research Team was used to calculate these ratings as that time the only Doordarshan channel was available. TRP is calculated by people meters and picture matching.

The purpose of this work is to evaluate the performance of TV show and also calculate the likes of a particular show or actors of that show and predicting popularity of that shows based on the text reviews. The reviews are getting on social networking websites like Facebook.
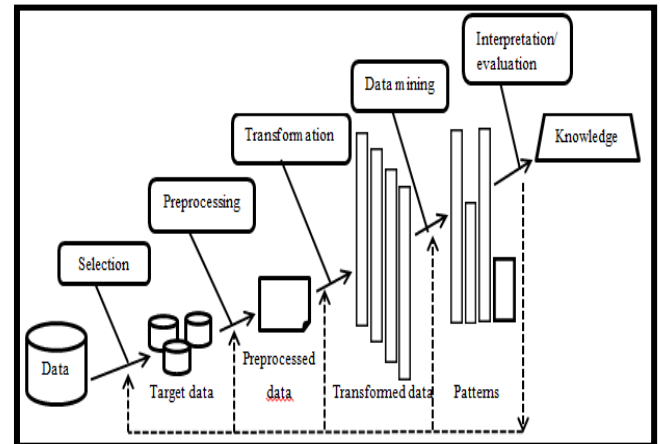
Figure 1 KDD Process [12]

II. RELATED RESEARCH WORK

Bria *et al*. (2007) [1] described various challenges and changing value chains in the mobile TV business. The main aim of this paper is to identify these challenges in the technology and industry strategy domain. Two technologies have been used in this study. First one is cellular network for delivery and second one is a separate broadcast network that is for mobile TV. The result of the study is that real need incentives for all actors in this ecosystem must be clear to successfully create "TV for mobile" adapted to mobile users. **Mhaisgawali and Giri (2014) [2]** discussed detailed descriptive and predictive analytics with TV ratings. The predictive analytics technique predicts that what will happen in future. In the end, the study predicted that any new mobile device can be launched in which city. **Rashid *et al*. (2013) [3]** discussed the areas, techniques and challenges of opinion mining. The result of this study is that now opinion mining trend is moving to the sentimental reviews of comments used in Facebook or twitter data. **Singh *et al*. (2011) [4]** discussed TRP as a measure of visual communication. The result of this study is that TRP of TV channel colors is at the top in Jammu city between 8pm and 11pm. **Sharma *et al.*(2013) [5]** discussed a review on data mining, its challenges, issues and applications. The result of the study is that data mining is very useful to businesses society government and also for an individual. It helps in strategy making, decision making and analysis.

Ramageri (2014) [6] discussed the various data mining techniques and applications. The result of this study is that data mining techniques and algorithm are helpful in finding the patterns to decide upon the future trends in businesses to grow. **Anand *et al.* (2018) [7]** discussed the two algorithms k-means and incremental k-means have been used for analysis of TV show popularity rating. In this paper comparison between the two algorithms is very clear on histograms. .They studied many algorithms but the complexity and efficiency is best in these algorithms. Using classification and clustering the best results achieved through incremental clustering algorithm at 97%.

III. METHODOLOGY

Random Tree has been developed by classification of data for prediction. During the time spent order WEKA 3.6.9 characterization apparatus has been utilized for arrangement process which uses distinctive classifiers, channels for ideal grouping, with the goal that choice can be made for information mining purposes. For numerical prediction the statistical methodology is used for regression analysis. For example in case of TRP what will be the views of people in future for any TV serial.

- Robustness is the ability of predictor to make correct predictions from noisy data.
- Scalability refers to construct the predictor efficiently.
- Accuracy of the predictor refers to how a predictor can guess the value of predictor attribute for a new data.

Algorithm

- *Finding the main factors:* The main factors have been finding out are responsible for determining the TRP of a channel or program. These factors are gender, age, qualification, marital status, occupation and area.

- *Determine the influencing factors:* Six major factors have been determined that influence the viewer's satisfaction. Data of these factors has been gathered from the historical data.

- *Analyse the viewer's satisfaction:* The impact of the factors have been analysed on the satisfaction of the viewer's.

- *Impact of TRP productive on viewer's satisfaction:* Impact of TRP productive has been analysed on viewer's satisfaction.

- *Investigate the gathered rating based reaction by classification tree:* Gathered reaction based on rating has been investigated by random tree in classification.

- *Predicting a viewer's opinion:* At the end analyse the viewer's opinion has been predicted.

```
┌─────────────────────────────────┐
│     Finding the main factors    │
└─────────────────────────────────┘
                 │
                 ▼
┌─────────────────────────────────┐
│  Determining the influencing    │
│  factors                        │
└─────────────────────────────────┘
                 │
                 ▼
┌─────────────────────────────────┐
│  Analyzing the viewer's         │
│  satisfaction                   │
└─────────────────────────────────┘
                 │
                 ▼
┌─────────────────────────────────┐
│  Impact of TRP productive on    │
│  viewer's satisfaction          │
└─────────────────────────────────┘
                 │
                 ▼
┌─────────────────────────────────┐
│  Investigate the gathered rating│
│  based reaction by classification│
└─────────────────────────────────┘
                 │
                 ▼
┌─────────────────────────────────┐
│  Predicting a viewer's opinion  │
└─────────────────────────────────┘
```
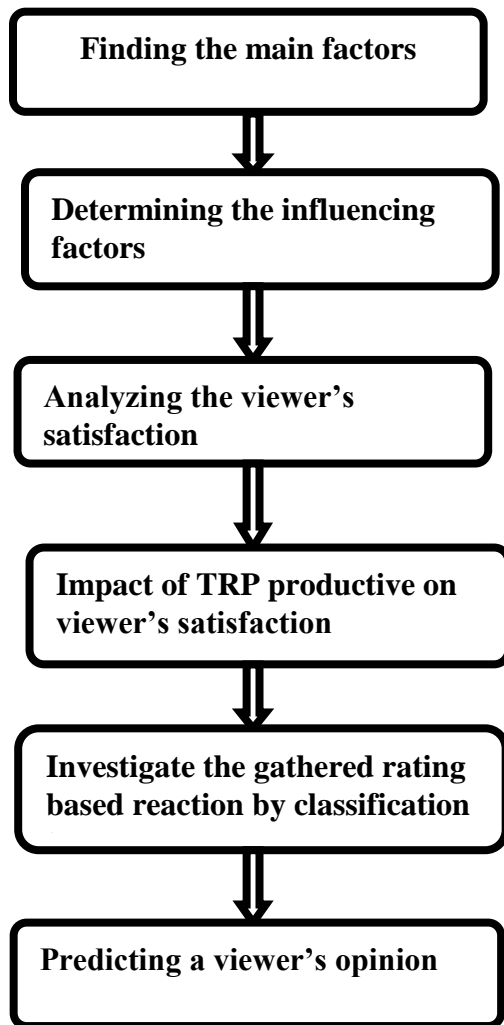
Figure 2 Flow chart of proposed work

Fig 2 represents flow of the proposed work that has been used for classification process. Random tree classifier is a collection of individual decision trees has been proposed to predict the TRP of television serials

4. RESULTS

Implementation of this work has been done in WEKA tool and WAMP server. WEKA tool is used for implement algorithms for classification, clustering, association rules, regression and data preprocessing. The other hand WAMP server is used for web designing and internal testing. It is also used to serve live websites. A process is followed and experimental results are designed to show the importance of television rating points in the field of television and film industry. We first examine the opinion of viewers on the basis of historical data. A prediction is created on the basis of historical data. In this process six genuine factors are examined. The data related to these factors has been taken from facebook.
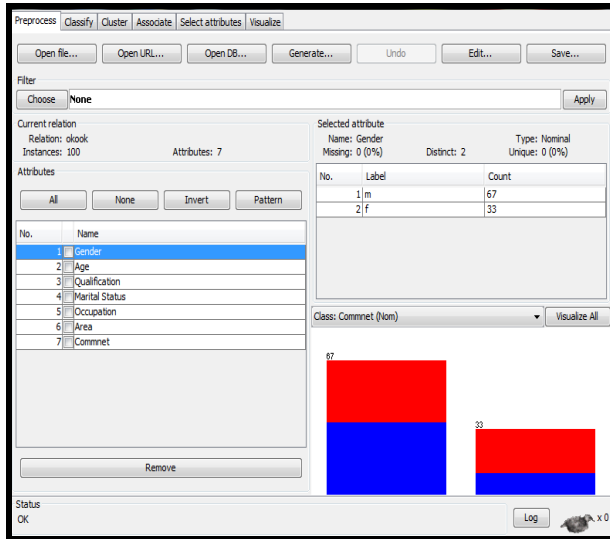
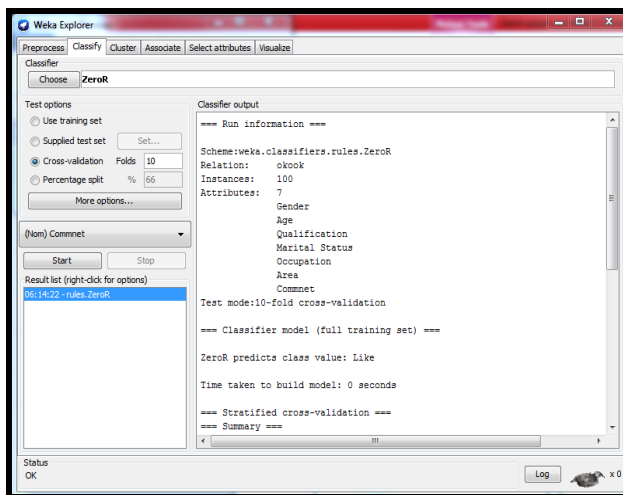Figure 3 Classification of males and females

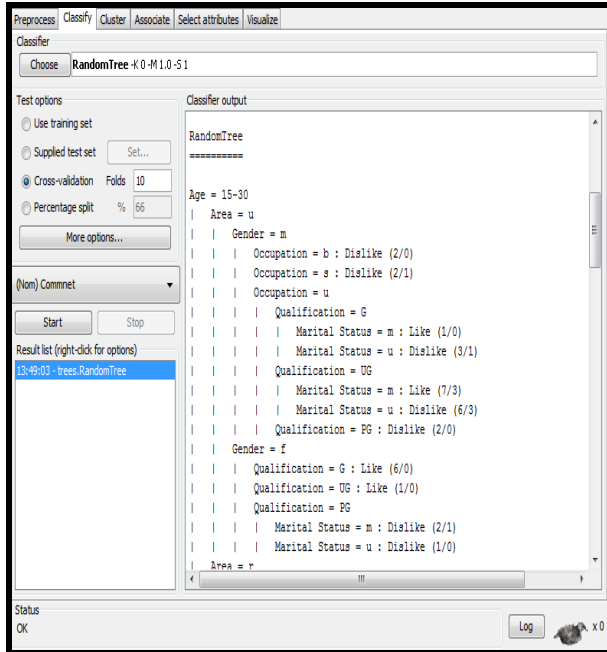Figure 4 Classification of 6 major factors

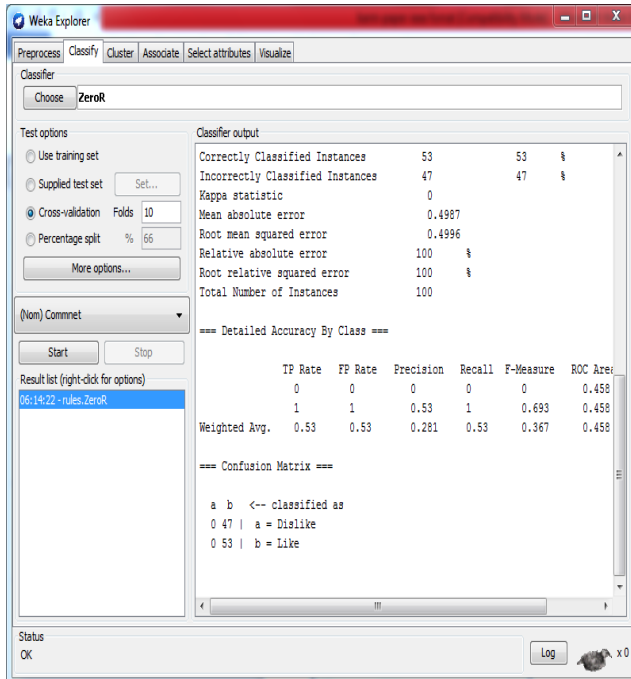Figure 5 Classification of historical data set in random tree

Figure 6 Classification of predicted data set using proposed   classifier

Fig 6 represents classification of historical dataset using proposed rule based classifier. This classifier generates support and confidence value from the dataset attributes. On the basis of these values a random tree has been generated by proposed approach.

Table 1 Accuracy tables for predicting TRP of different serials

Different serials	Matched in %	Unmatched in %	Accuracy in %
TU SURAJ MEIN SANJH PIYA JI	80	20	80
YEH RISHTA KYA KEHLATA HAI	37.5	62.5	37.5
TARAK MEHTA KA OOLTAH CHASHMA	25	75	25

Table 1 represents various serials prediction is matched and unmatched have been evaluated for performance evaluation of proposed work. These parameters are classification model that classified different instances into different classification on the basis of classification rules. This table contains total number of matched and unmatched classified instance out of 100 instances in testing set.
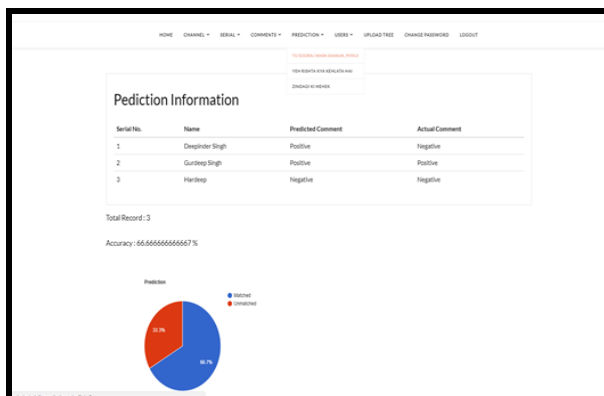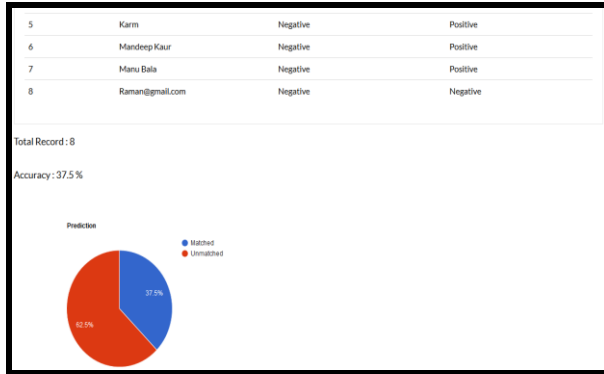

Figure 7 Prediction of "TU SURAJ MEIN SANJH PIYA JI"

5	Karm	Negative	Positive
6	Mandeep Kaur	Negative	Positive
7	Manu Bala	Negative	Positive
8	Raman@gmail.com	Negative	Negative

Total Record : 8

Accuracy : 37.5 %

Figure 8 Prediction of "YEH RISHTA KYA KEHLATA HAI"

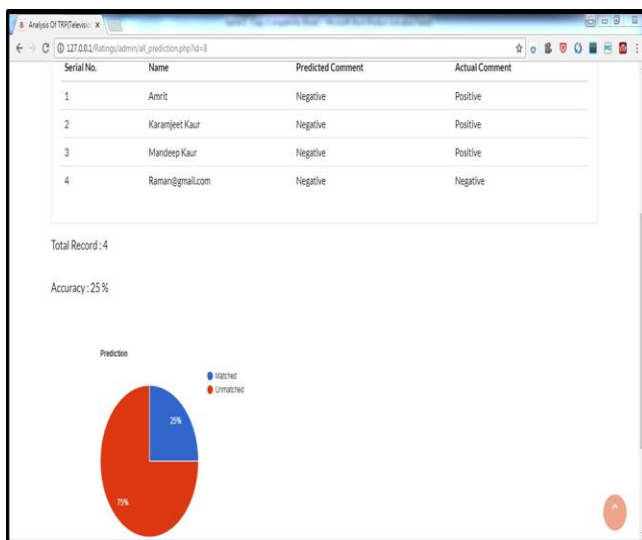| Serial No. | Name | Predicted Comment | Actual Comment |
1	Amrit	Negative	Positive
2	Karamjeet Kaur	Negative	Positive
3	Mandeep Kaur	Negative	Positive
4	Raman@gmail.com	Negative	Negative

Total Record : 4

Accuracy : 25 %

Figure 9 Prediction of "TARAK MEHTA KA OOLTAH CHASHMA"

Fig 7, 8 and 9 represents graphical representation of the accuracy achieved by different classifiers using classification on historical dataset.

Table 2 Performance evaluation parameters for PREDICTION of serials

Approaches	Precision	Recall	F-measure	ROC	TP rate	FP rate
Correctly classified	0.464	0.553	0.505	0.464	0.553	0.566
Incorrectly classified	0.523	0.434	0.474	0.464	0.434	0.447
Weighted average	0.495	0.49	0.489	0.464	0.49	0.503

Table 2 represents various parameters that have been computed for performance evaluation of proposed work. These parameters are precision, recall, f-measure, Roc, TP rate and FP rate. On the basis of these parameters various classification algorithms have been validation for computation of best classifier based on rules that can be used for classification and prediction.

CONCLUSION

In the thesis, the historical data has been analyzed to obtain trends in the viewership. By the help of historical data the future of the television serials are predicted. The classification tree helps to classify the historical demographic data & obtain the predicted future viewers. The WEKA tool was used to analyze the whole process to ensure the viewer's choice for future. The programming language PHP was used to implement the webpage for TRP representation.

FUTURE SCOPE

● Prediction has been created on the basis of historical data so in future the real time data means primary data can be used for the prediction.

● In this 6 major factors were consider so more factors like contact no. and security question can be added.

● On behalf of these factors like & dislike conditions were created, in future comments are also applied.

● Random tree was created on the basis of historical data; the other tree like NB tree can be applied for it.

● WEKA tool was used for classification; in future other data mining tool like Hadoop and Hive can also try.

REFRENCES

[1]. A. Bria, P. Karrberg, P. Andersson(2007), **"TV in the mobile or TV for the mobile: challenges and changing value chains"**, IEEE International Symposium on Personal, Indoor and Mobile Radio Communications, pp. 1-5.

[2]. A. Mhaisgawali, N. Giri (2014), **"Detailed descriptive and predictive analytics with twitter based TV ratings"**, International Journal of Computing and Technology, **1(4)**, pp. 125-130.

[3]. A. Rashid, N. Anwer, M. Iqbal, M. Sher (2013), **"A survey paper: areas, techniques and challenges of opinion mining"**, International Journal of Computer Science Issues, **10(2)**, pp. 18-31.

[4]. A. Singh, S. K. Mehta, H. G. Mishra (2011),**"TRP as a measure of visual communication: a study of Jammu city, India"**, International Conference on Economics and Finance Research, **2(7)**, pp. 123-137.

[5]. B. R. Sharma, D. Kaur, M (2013) **"A review on data mining: its challenges, issues and applications"**, International Journal of Current Engineering and Technology, **3(2)**, pp. 695-700.

[6]. B. M. Ramageri (2014), **"Data mining techniques and applications"**, Indian Journal of Computer Science and Engineering, **1(4)**, pp. 301-305.

[7].   D. Anand, A. V. Satyavani, B. Raveena, M. Poojitha (2018), **"Analysis and prediction of Stelevision show popularity rating using incremental k-means algorithm"**, International Journal of Mechanical Engineering and Technology, **9(1)**, pp. 482-489.

[8].   E. Panova, A. Raikov O.Smirnova(2015), **"Cognitive television viewer rating"**, International Young Scientists Conference on Computational Science, **66**, pp. 328–335.

[9].   G. Vinodhini, R. M. Chandrasekaran (2012), **"Sentiment analysis and opinion mining: a survey"**, International Journal of Advanced Research in Computer Science and Software Engineering, **2(6)**, pp. 283-292.

[10].   L. Molteni, J. P. D. Leon (2016), **"Forecasting with twitter data: an application to USA TV series audience"**, International Journal of Design and Nature and Ecodynamics, **11(3)**, pp. 220-229.

[11].   N. Rikhi (2015), **"Data mining and knowledge discovery in database"**, International Journal of Engineering and Trends Technology, **23(2)**, pp. 64-70.

[12].   P. G. Malur, N.Lakshmikantha, V.Prashanth(2014),**"Reeling the reality: a study on contemporary reality shows and their Influence on other entertainment program genres"**, International Research Journal of Social Sciences,**3(8)**, pp. 35-38.