ISSN 2348-1218 (print)
International Journal of Interdisciplinary Research and Innovations ISSN 2348-1226 (online)
Vol. 3, Issue 2, pp: (34-40), Month: April - June 2015, Available at: www.researchpublish.com

Movie Rating System Based on Opinion Mining

Zainab Mirza[1], Mehwash Khan[2], Saima Khan[3], Khurshid Khatri[4]

[1,2,3,4,]M.H Saboo Siddik College of Engineering, Mumbai, India

Abstract: **Opinion Mining (also referred as Sentiment Analysis) refers to the use of natural language processing, text analysis and computational linguistics to identify and extract subjective information in source materials. For any type of information gathering, we always see what the opinions of people about that product are or any service provided. With the growing availability and popularity of opinion-rich resources such as online review sites and personal blogs, new opportunities and challenges are now available for people. Opinion Mining is such a field that helps user know the reviews about any product or service he / she is interested in. Opinion Mining is a task of extracting from a set of documents expressed by a source on a specified target.**

Keywords: **Opinion Mining, Sentiment Analysis, Sentiment, Naive Bayes.**

I. INTRODUCTION

The World Wide Web is increasing at an alarming rate not only in size but also in the types of services and contents provided. With the rapid growth of e-commerce, more and more products are sold on the web and more number of customers is also buying products online. Now-a-days people are dependent on internet be it any information about product or services or reviews for it. Due to the large number of reviews, it is hard for potential customer to get efficient review details, it is a challenging task. Also instead of referring the newspapers that only contain a five star based rating which are usually not accurate, therefore, many users are dependent on websites that provides ratings as well as a brief review about others opinion.

Opinion Mining is an area of text mining that has recently received a lot of attention due to huge amount of opinion or reviews available in web documents. Thus Opinion mining is also called as Sentiment Analysis. Opinion mining can be defined as a sub-discipline of computational linguistics that focuses on extracting people's opinion from the web. Opinion mining typically occurs in two or three stages:

1. The input text is split into sections, such as sentences, and each section tested to see if it contains any sentiment: if it is subjective or objective.
2. The subjective sentences are analyzed to detect their sentiment polarity.
3. The object about which the opinion is expressed may be extracted.

Our project uses concept of Opinion Mining which helps to extract user's opinions in the form of comments. It gives results based on the polarity of the comments i.e., positive or negative polarity.

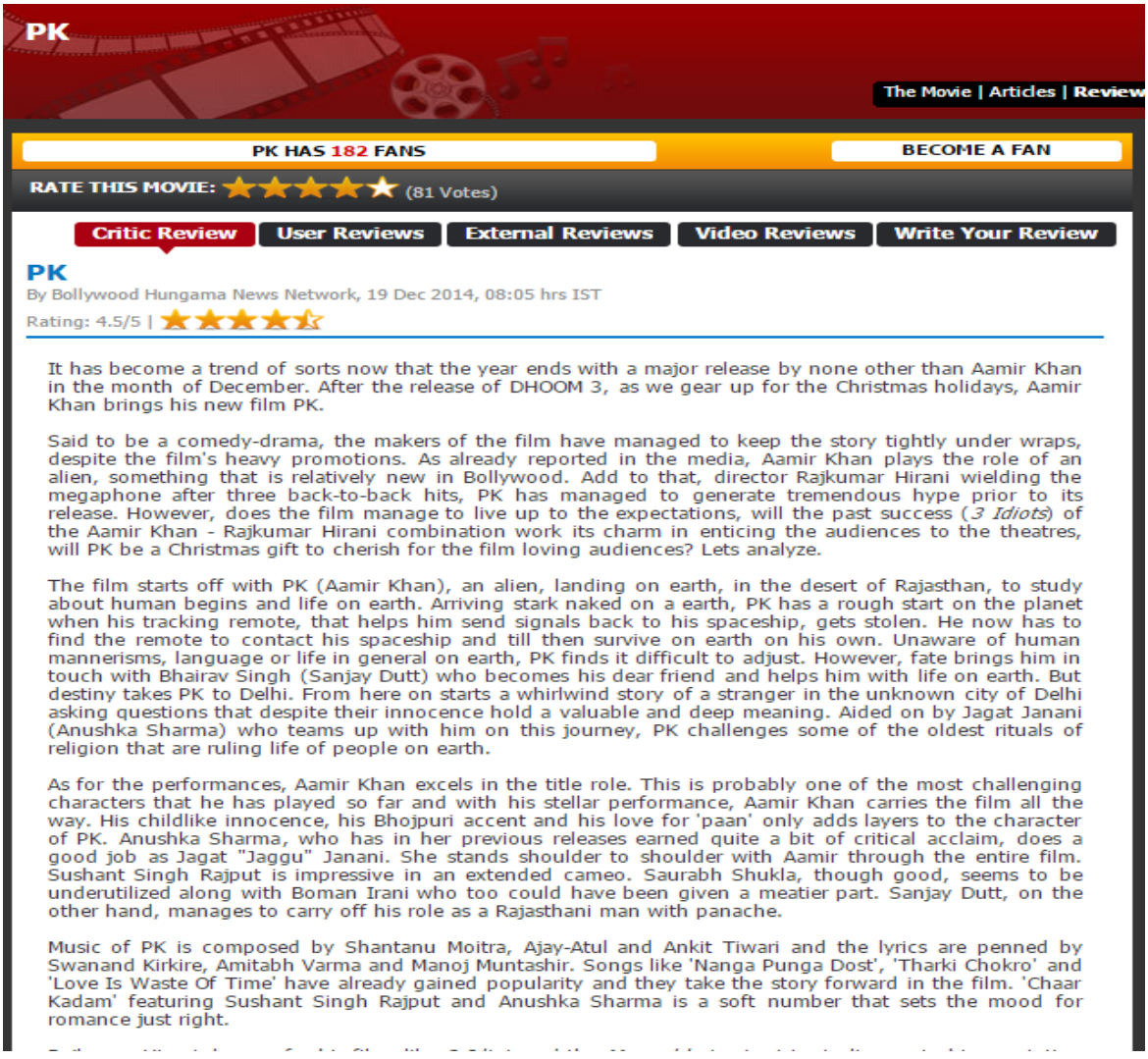
Based on the comments given by the user, we classify the movie in the following three categories:

1. Not Worth Watching
2. 50-50
3. Worth Watching

II. RELATED WORK

Most of the movie websites categorizes movie on the basis of rating provided by the user. In our project, movie is classified on the basis of comments given by the user.

ISSN 2348-1218 (print)
International Journal of Interdisciplinary Research and Innovations ISSN 2348-1226 (online)
Vol. 3, Issue 2, pp: (34-40), Month: April - June 2015, Available at: www.researchpublish.com



ISSN 2348-1218 (print)
International Journal of Interdisciplinary Research and Innovations ISSN 2348-1226 (online)
Vol. 3, Issue 2, pp: (34-40), Month: April - June 2015, Available at: www.researchpublish.com

III. WORK FLOW

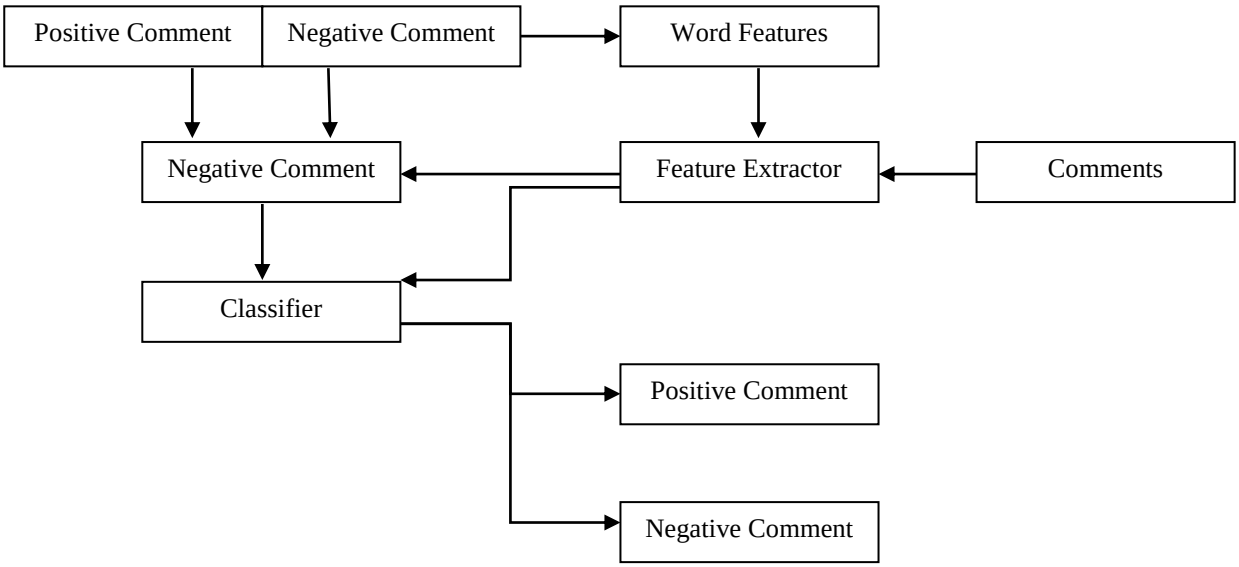


Fig: Workflow Process

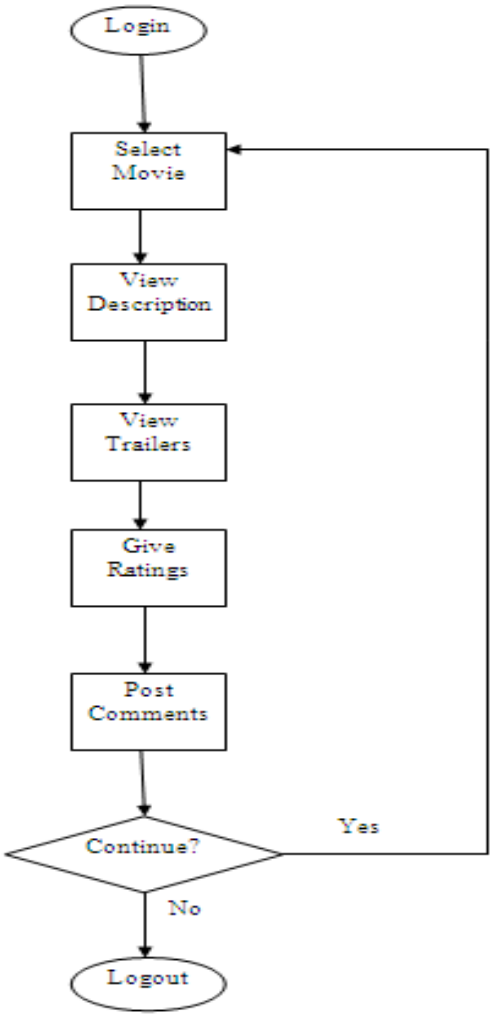


Fig: Flow of our project

ISSN 2348-1218 (print)
International Journal of Interdisciplinary Research and Innovations ISSN 2348-1226 (online)
Vol. 3, Issue 2, pp: (34-40), Month: April - June 2015, Available at: www.researchpublish.com

IV. WORKING

Our proposed System will have two types of users, registered user or unregistered user. Unregistered user can view the trailer and can only read comments. The registered user can view trailers give comments and rating.

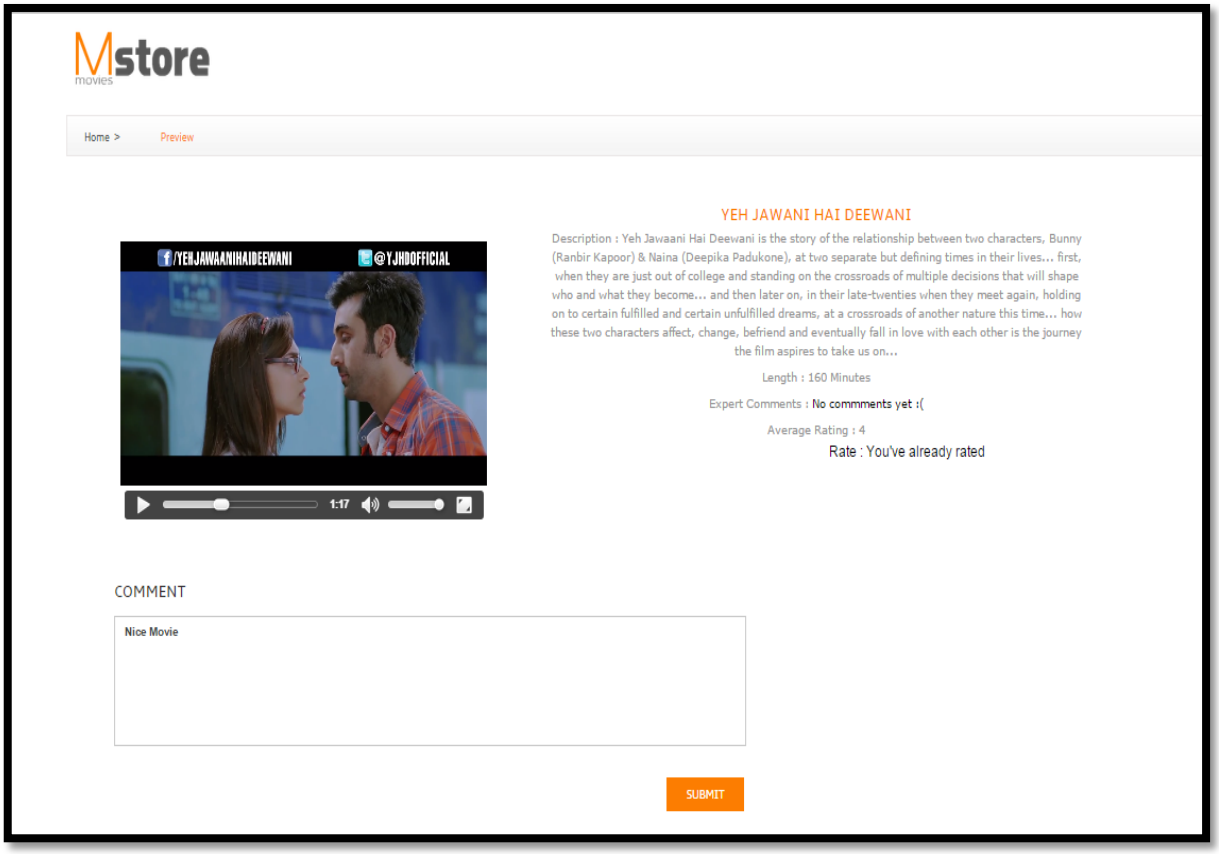


Fig. Trailer, Comments And Rating

The privilege of being a registered user is that he gets suggestions based on his search history and the movie he/ she searches are visible in recently viewed section.

Fig. Suggestions

ISSN 2348-1218 (print)
International Journal of Interdisciplinary Research and Innovations ISSN 2348-1226 (online)
Vol. 3, Issue 2, pp: (34-40), Month: April - June 2015, Available at: www.researchpublish.com

The comments given by user are stored in database and then keywords are extracted. The extracted keywords are checked for its polarity (i.e. positive and negative) using Naive Bayes Algorithm

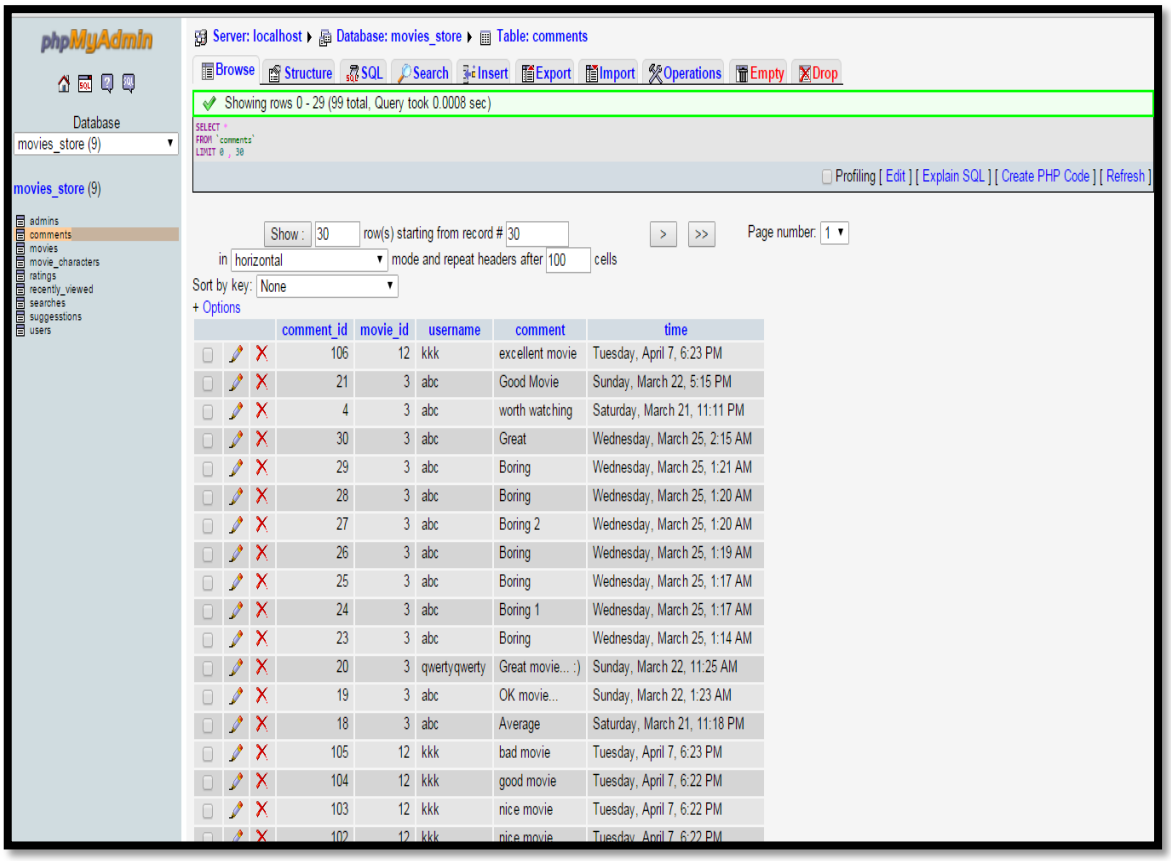


Fig. Comments Database

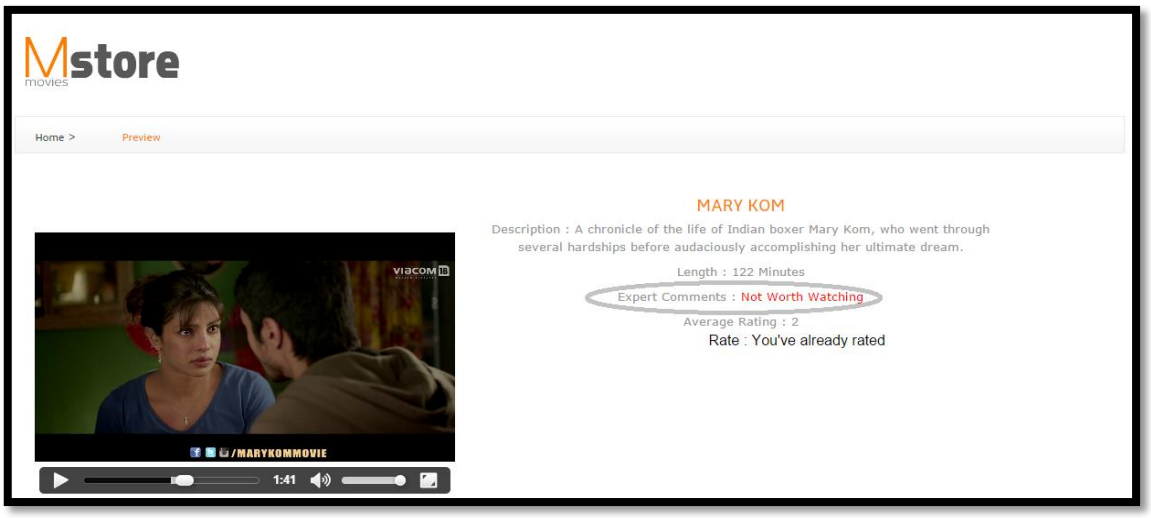


Fig. Expert Comment view

The comments provided by registered user are classified in the following three categories:

- Not Worth Watching
- 50-50
- Worth Watching

ISSN 2348-1218 (print)
International Journal of Interdisciplinary Research and Innovations ISSN 2348-1226 (online)
Vol. 3, Issue 2, pp: (34-40), Month: April - June 2015, Available at: www.researchpublish.com

V. BAYES ALGORITHM

As written in [4], Bayes Theorem is a statement from probability theory that allows for the calculation of certain conditional probabilities. Conditional probabilities are those probabilities that reflect the influence of one event on the probability of another event. The term generally used in Bayes' theorem is prior probability and posterior probability. The prior probability of a hypothesis or event is the original probability obtained before any additional information is obtained. The posterior probability is the revised probability of the hypothesis using some additional information or evidence obtained.

Bayes' Theorem can be written as:

$$P(A|B) = \frac{P(B|A)\,P(A)}{P(B)} \qquad \text{(Eq.1)}$$

Where,

P (A) is the prior probability of A

P (B) is the prior probability of B

P (A|B) is the posterior probability of A given B

P (B|A) is the posterior probability of B given A

Since the denominator P(B)in Eq. 1 is the probability of the evidence without any knowledge of the event A, and since the hypothesis A can be true or false, Bayes theorem can also be written as,

$$P(A|B) = \frac{P(B|A)\,P(A)}{P(B|A) \times P(A) + P(B|\neg A) \times P(\neg A)} \qquad \text{(Eq. 2)}$$

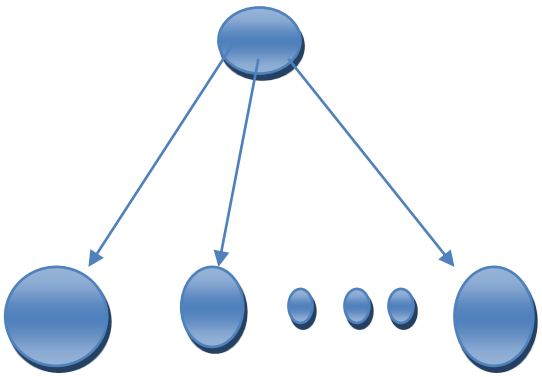
Where,

P (¬A) is the probability of A being false

P (B|¬A) is the probability of B given A is false

NAÏVE BAYESIAN:

A naïve Bayes classifier is a probabilistic classifier based on applying Bayes' theorem with strong independence assumptions. The naïve Bayesian classifier was first described in 1973 and then in 1992. When represented as a Bayesian network, a naïve Bayes classifier has the structure depicted in Figure 1. It shows the independence assumption among all features in a data instance.



ALGORITHM:

A naïve Bayes classifier can be defined as below. Variables are denoted using capital letters such as X_i, and their values will be denoted by lower-case letters such as X_i, and sets of variables are denoted by boldface letters such as X.

Let $\mathbf{X} = \{X_1, \ldots, X_n\}$ be a finite set of observed random variables, called features, where each feature takes values from its domain D_i. The set of all feature sets is denoted by $\Omega = D_1 \times \ldots \times D_n$. Let C, such that $c \in \{0, \ldots, u-1\}$, be an unobserved random variable denoting the class of a set of features.

ISSN 2348-1218 (print)
International Journal of Interdisciplinary Research and Innovations ISSN 2348-1226 (online)
Vol. 3, Issue 2, pp: (34-40), Month: April - June 2015, Available at: www.researchpublish.com

A hypothesis $h : \Omega \rightarrow \{0, \dots, u-1\}$, that assigns a class to any given set of variables is defined as a classifier. Each class c is assigned a discriminant function $f_c(\boldsymbol{x})$, $c = 0, \dots, u-1$. The classifier selects the class with the maximum discriminant function on a given set of variables, written as $h(\boldsymbol{x}) = \arg max_{c \in \{0,\dots,-1\}} f_c(x)$.

The Bayes classifier $h^*(\boldsymbol{x})$ uses the posterior probabilities given a set of variables as the discriminant function, i.e. $f^*(\boldsymbol{x}) = (C = c \mid \boldsymbol{X} = \boldsymbol{x})$. Applying Bayes' theorem from Eq. 1 to this function gives $(C = c \mid \boldsymbol{X} = \boldsymbol{x}) = \underline{(C = c \mid X = x)\ P(C = c)}$. Since P(X = x) is the same for all

P(X = x)

classes it can be ignored. Hence, the Bayes' discriminant function can be written as

$f^*(\boldsymbol{x}) = P(\boldsymbol{X} = \boldsymbol{x} \mid C = c)\ P(C = c)$

where $P(\boldsymbol{X} = \boldsymbol{x} \mid C = c)\ P(C = c)$ is called the 11 class-conditional probability distribution (CPD). Thus the Bayes' classifier written as in Eq. 3 finds the maximum posterior probability hypothesis given x.

$h^*(\boldsymbol{x}) = \arg max_c (\boldsymbol{X} = \boldsymbol{x} \mid C = c)\ (C = c)$ (Eq. 3)

Applying the assumption that features are independent given the class on Eq. 3, we can get the naïve Bayes classifier.

$f_c^{NB}(\boldsymbol{x}) = \prod_{j-1}^{n} (X_j = x_j \mid C = c)\ (C = c)$ (Eq. 4)

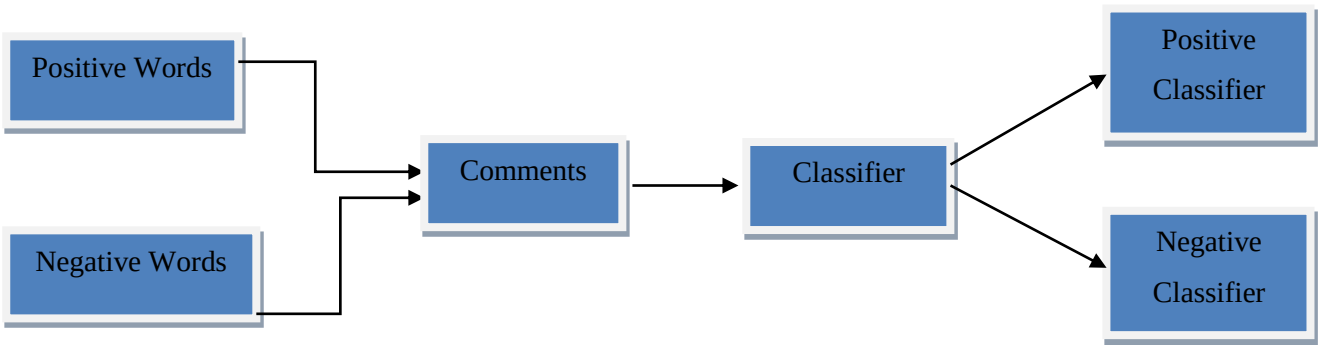


Fig. Algorithm Of Naive Bayes

VI. CONCLUSION

There are many advantages of using opinion mining like in many e-commerce website or blogs wherein the user gets reviews of any product or service. Opinion mining is a very vast concept and it is the base of our project. We have used the concept of opinion mining in our project on the comments given by the user. Use of Naïve Bayes Algorithm decides the polarity of the comments with the help of which expert comments are provided. The unique thing of the project is that it extracts the user comments and automatically generates expert comments. Unlike the unregistered user, the registered user gets benefit of suggestions which will be generated by the system. The suggestion is on the basis of actor and actress the user frequently searched for.

REFERENCES

[1] Alexandra Balahur, Ralf Steinberger, Erik van der Goot, Bruno Pouliquen, Mijail Kabadjov: Opinion Mining on Newspaper Quotations

[2] Periakaruppan Sudhakaran , Shanmugasundaram Hariharan and Joan Lu Research Directions, challenges and Issues in Opinion Mining

[3] 2sentiment_a_v0 {PPT}.

[4] Bandana Garg, Design and Development of Naïve Bayes Classifier.