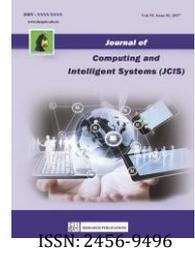Original Research Article

Journal of Computing and Intelligent Systems

Journal homepage: www.shcpub.edu.in

Sentiment Analysis Using Online Videos Reviews

A Priyanga #1, J Roserin Thiravia #2

Abstract — One of the most important functions of NLP (natural language processing) is sentiment analysis. This research are capturing , sentiment contradiction categorization method for sentiment analysis contradiction is preferred with brief process illustration. We are using data from online video reviews obtained from YouTube. In this work perform review level designate demonstrations are completed with a positive outcome. By crawling YouTube data with the keywords "sentiment analysis" "online video reviews" "natural language processing". We utilized the random forest algorithm and the naïve bayes algorithm to do sentiment analysis. We are manually crawl the site using an access token obtained from YouTube and rapid miner tools to collect the needed data and information. Given random forest when compared with the naïve bayes, the result of emotion measurement is 86 % accuracy.

Keywords - Natural language processing, Contradiction, Categorization, Random Forest, Naïve Bayes, Sentiment analysis.

I. INTRODUCTION

Big Sentiment is a feeling-driven attitude. Opinion mining is also called sentiment analysis. The computational analysis of attitudes, feelings, views, news, comments, and feedback is known as sentiment analysis.

Internet is a great resource. Users to publish their emotions on numerous social media specifying whether the view is favorable, negative, or neutral. Therefore Datasets used in the work are both true and publicly available. Data utilized in this document is a collection of internet video reviews gathered from YouTube. Before being posted, each video review is inspected. Online video reviews were collected and analyzed to determine the highest and lowest rating levels for various categories of videos

Fig :1 Rating system for YouTube videos

II. RELATED WORKS

Sentiment analysis often known as sentiment information, the internet is a valuable resource [1]. API developers can gather status data and user information . Review material, and streaming API can allowing them to design applications. with the help of enormous web data.[3-5]. The technique of assessing reviews known as opinion mining. Emotion analysis is currently frequently employed in sentiment analysis a part of computer science, are widely used to classify people's emotions [6,7]. There are various weaknesses in online data that could stymie the process. Analysis of public opinion ,while anyone publish their content. Some spam is completely worthless, while others contain irrelevant or fake opinions[8].We will simply examine some past work in this section, on which our research is mostly based. The parts of speech refer to the syntactic duties. In English, there are eight parts of speech: verb, pronoun, noun, adjective, adverb, and preposition. Part-of-speech (POS) taggers are used in natural language processing[9,10]. There are 1.6 million machine-tagged tweets in the corpus. Each message is labeled according to the feelings it evokes, both positive and negative this study addresses a key issue in sentiment analysis: sentiment polarity categorization using an algorithm [11,12].

Opinion mining is a technique for determining the categorization of naïve bayes models. It's used to demonstrate the accuracy of categorization effects[13]. Information are used to defect substantially overcome [14]. In analysis of public opinion emotion level. The entity focuses on what people enjoy and dislike based on their feedback[15,16]. Using Twitter data, Gann and colleagues identified 6,747 tokens, each of which is assigned a sentiment score is used to compute it[17].

* Corresponding author: E-mail: 1rithupriya16@gmail.com, 2gthiraviajohn1999@gmail.com

1 &2 Research Scholar, Department of Computer Science, Dr. Umayal Ramanathan College for Women, Karaikudi , Tamilnadu, India.

III. METHODOLOGY
A. DATA COLLECTION

The Information used in this study is based on a collection from YouTube video reviews. 2.1 million online video reviews data was gathered in total, with the videos falling into four broad categories: instructional, humor, cookery, and storytelling. More than 1.3 million people watched those internet video reviews. The following information is included in each review: 1.the reviewer's ID; 2.the channel's name; 3.the rating; and 4.the review text.
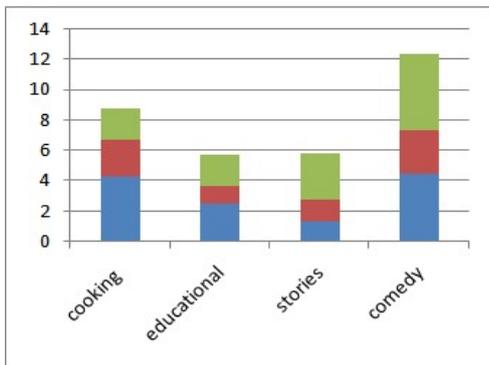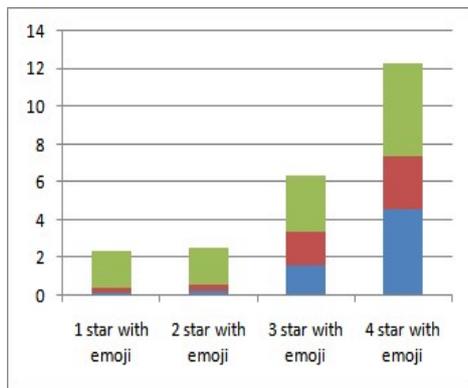
Fig 2: Video categories

Fig :3 Review categories.

B. SENTIMENT ANALYSIS

Sentiment Analysis terms Subjective information in an expression. Expressions can be classified as Positive, Negative or Neutral. The accuracy of a sentiment analysis system is, in principle, how well it agrees with human judgments. This is usually measured by variant measures based on precision and recall over the two target categories of negative and positive texts.

C. NAVIE BAYES CLASSIFIER

Naïve Bayes classifier is high-bias and low-variance classifier that also known as simple bayes or independence bayes. The most common use of the naïve bayes classifier is sentiment analysis.

probabilistic classifier, the naïve bayes algorithm This approach is excellent for making forecasts and forecasting data based on historical outcomes.

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

A and B are events and $P(B) \neq 0$. $P(A)$ is the **Prior** of A. $P(A|B)$ is a posteriori probability of B.

D. RANDOM FOREST

The Random Forest Algorithm is another prominent Big Data Analytics approach. Random forest is a big data Model. The variety of samples, using the average for regression and the majority vote for classification. "Ensemble" simply refers to the combination of several nodes. Cross validation ensures a higher level of accuracy. It increases the accuracy of a decision tree by reducing over fitting concerns and decreasing variance. The Constructed Classification Structure can also be modified. The dataset was separated into various subsets in this algorithm, which were then expanded with nodes.

The final leaf node value is then taken into account for prediction. The average of all trees, based on the total key features in an RF.

$$RF\ fi_i \quad = \quad \frac{\Sigma_{j\epsilon all\ tree}\ ^{norm\ fi}{}_{ij}}{T}$$

The RFi_i represents the feature's importance, the norm fi sub (ij) represents the normalized importance T is the total number of trees, and I in tree j.

IV. RESULT AND DISCUSSION

In YouTube videos, we employ the keyword "Sentiment analysis" . The naive bayes algorithm and the random forest method are used to analyze, and interpret all accessible research with the goal of delivering answers. At the time, net videos were average, thus a neutral opinion was formed. In this paper, the random forest achieves the highest level of accuracy when compared to nave bayes
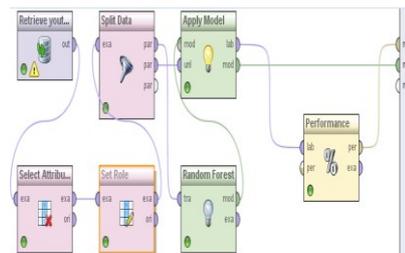
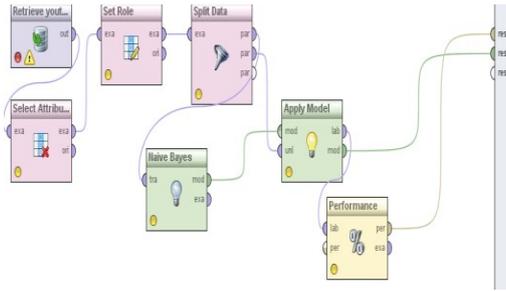Fig 4:data crawling process in random forest .

Fig 5: data crawling process in naïve bayes

Fig 4 and 5 The rapid miner crawling process is depicted. It is used to link data by entering an access code obtained from YouTube. The next operation is to pick attributes, which are subsequently used to get properties like username and text. The information is kept in an excel spreadsheet. For data extraction, use the rapid miner tool.

.

Fig 6: Sentiment analysis for random forest

Fig 7: sentiment analysis for naïve bayes

Figure 6, 7 depicts the outcome of sentiment analysis, which included over 3.3 thousand negative comments, over 5.4 thousand good comments, and the remaining 5.4 thousand neutral remarks. As a result of these work, comedic videos receive more views and favorable feedback. As a result, the majority of people are watching comedic videos.
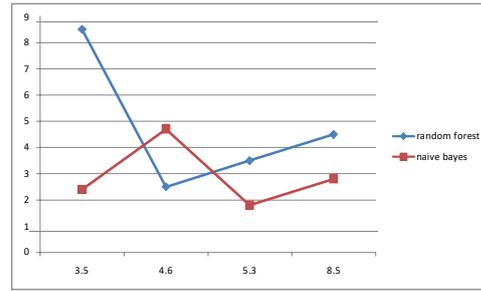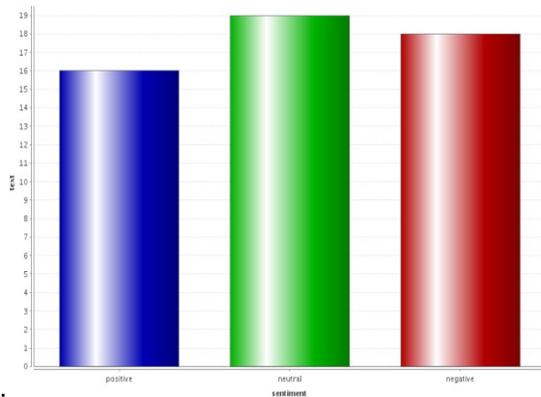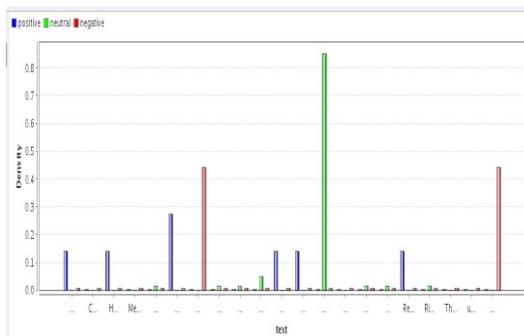

Fig 8: review level categorization

Fig 8 the Positive, neutral, and negative data are regularly divided into three categories. The naive bayes and random forest may be noticed. When compared to naive bayes, random forests are often superior.

Fig 9 . Performance Vector of Random Forest.

Fig 9 & 10 shows Accuracy, Classification Error, Confusion Matrix, Recall, Precision, Relative Error, Absolute Error value for Decision Tree and Random Forest Algorithm. The ability of an instrument to measure the accurate value is known as accuracy. Accuracy is obtained by taking small readings. The closeness of two or more measurements to each other is known as the precision . Recall is the fraction of relevant instances that were retrieved. A measure that combines precision and recall is the harmonic mean of precision and recall, the traditional F-measure or balanced F-score. The relative error is ,ratio of the absolute error of the measurement to the actual measurement. Absolute error is, difference between measured or inferred value and the actual value of a quantity.

```
accuracy: 97.30%
ConfusionMatrix:
True:   positive      neutral negative
positive:      4         0        0
neutral:       1        31        0
negative:      0         0        1
classification_error: 2.70%
ConfusionMatrix:
True:   positive      neutral negative
positive:      4         0        0
neutral:       1        31        0
negative:      0         0        1
weighted_mean_recall: 93.33%, weights: 1, 1, 1
ConfusionMatrix:
True:   positive      neutral negative
positive:      4         0        0
neutral:       1        31        0
negative:      0         0        1
weighted_mean_precision: 98.96%, weights: 1, 1, 1
ConfusionMatrix:
True:   positive      neutral negative
positive:      4         0        0
neutral:       1        31        0
negative:      0         0        1
absolute_error: 0.057 +/- 0.178
relative_error: 5.72% +/- 17.76%
```

Fig 10 . Performance Vector of Navie Bayes.

IV. CONCLUSION

Naïve bayes and random forest algorithms on YouTube data with the keyword "sentiment analysis" during the first and second weeks of December 2021. During that time span, the study revealed 36% positive sentiment, 54% neutral sentiment, and 41% negative emotion. People's attitudes toward inter an accuracy of 86 percent. Random Forest Produce (78%) and Navie Bayes produce (97%) . As a result, the Navie Bayes gave precise results in the Prediction Examination.

ACKNOWLEDGEMENT

REFERENCES

[1] Brenda A. Januario Sentiment Analysis Applied to News from the Brazilian Stock Market (2022).

[2] Chien-Cheng Lee , Zhongjian Gao , Chun-Li Tsai BERT-Based Stock Market Sentiment Analysis (2020).

[3] Sentiment Analysis for Social Media using SVM Classifier of Machine Learning(2019).

[4] Rajkumar Sujatha , Varad Abhyankar , Abhishek Gehlot , Priyank Gupta , Sivakumar Subramaniam Stock Market Trend Prediction Using Regression Model, RNNs, and Sentiment Analysis (2021).

[5] Pak A, paroubek p(2021)Twitter as a corpus for sentiment analysis using subjectivity summarization based on minimum cuts,: proceedings of the 42Nd Annual meeting association for computational linguistics,ACL'02.association for computational linguistics,Stroudsburg,PA,USA,pp417-424

[6] R.Aulianita,L.Utami, N.Musyaffa, G.Wijaya, A.Mukhayaroh, andA.Yoraeni,Sentiment analysis Review of smartphones with Artificial Inteligent Camera Technology Using Naïve Bayes and n-gram character selection. J.phys. conf.ser.vol 1641,p.012076,Nov.2020,doi.10.1088/1742-6596/1/012076.

[7] T.mustaqim,K.umam,and M.A.Muslim, Twitter text mining for sentiment analysis on government's response to forest fires with vaderlexcon polarity detection and k-nearest neighbor algorithm , j.phys.conf.ser.,vol.1567,p. 032024,jun.2020,doi:10.1088/1742-6596/1567/3/032024.

[8] Shilpa B L , Shambhavi B R Combined deep learning classifiers for stock market prediction: integrating stock price and news sentiments (2021).

[9] P Padhanarath , Y Aunhathaweesup , S Kiattisin Sentiment analysis and relationship between social media and stock market: pantip.com and SET(2019).

[10] Role of Social Sentiment Analysis in Stock Trends Forecasting (2019).

[11] Zhaoxia Wang , Zhenda HU , Fang LI , Seng-Beng HO Learning-Based Stock Market Trending Analysis by Incorporating Social Media Sentiment Analysis.

[12] Sentiment Analysis on Social Media Big Data With Multiple Tweet Words (2019).

[13] R.Watrianthos , S.Suryadi, D.Irmayani, M.Nasution, and E.F.S.Simanjorang, sentiment analysis of traveloka app using naïve bayes classifier method,Int.J.Sci.Technol.Res vol.8, no.7,pp.786-788,2019,doi:10.31227/osf.io/2dbe4.

[14] Information available Uniform Resource Locator is "www.amazon.com".

[15] Erick Omuya , George Okeyo , Michael Kimwele Sentiment Analysis on Social Media using Machine Learning Approach (2021).

[16] Amit Purohit Sentiment Analysis of Customer Product Reviews using deep Learning and Compare with other Machine Learning Techniques (2021).

[17] Nimesh V Patel , Hitesh Chhinkaniwala Investigating Machine Learning Techniques for User Sentiment Analysis (2019).