IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

FAKE REVIEW DETECTION USING MACHINE LEARNING TECHNIQUES

Sangamesh Gama[1], Abhinandan V[2], Aishwarya C A[3], Arshiya Sultana[4]

[1]Assistant Professor, [2]Student, [3]Student, [4]Student
Department of Information Science and Engineering,
Atria Institute of Technology, Bangalore, Karnataka, India

Abstract: Online reviews play a vital role in today's business and commerce. In the world of e-commerce, reviews are the best signs of success and failure. Business that has good reviews get a lot of free exposure on websites and pages that have good reviews show up at top of the search results. Fake Reviews are everywhere online. Online fake reviews are the reviews which are written by someone who has not actually used the product or the services. Because of the cut-throat competitions, sellers are now willing to resort unfair means of making their product shine out .This work introduces some supervised machine learning techniques to detect fake online reviews and also be able to block the malicious users who post such reviews.

Index Terms **- Fake Review Prediction, Machine Learning, Naïve Bayes Classifier, Natural Language Processing (NLP)**

I. INTRODUCTION

Many researches have been studying on the detection of these fake online reviews. Some approaches are review content based and some are based on behaviour of the user who is posting reviews. Content based study focuses on what text of the review where user behaviour based method focuses on country, IP-address, number of posts of the reviewer etc. Most of the proposed approaches are supervised classification models. Few researchers, also have worked with semi-supervised models. Semi-supervised methods are being introduced for lack of reliable labelling of the review.

Technologies are changing every day. The new Technologies are helping people to do their work efficiently. Online marketplace is one of the form of new technologies. People can shop online using online websites. Since people shop online, almost every person tend to look at reviews before making buying decision .Online Reviews are not just reviews they are ultimate form of Advertisement to the companies. Reviews also impact on the reputations of the companies. With the Spread of online marketplace, Fake reviews are becoming great matter of concern. People can use unfair means to make their product shine. Also, competitive companies can try to damage each other's reputation by providing fake negative reviews. In this paper we make use of machine learning Techniques like Naïve Bayes Classifier and also Natural Processing Language to detect and to find the respective results.

II. RELATED WORK

Ott et al. used three techniques to perform classification. These three techniques are- genre identification, detection of psycholinguistic deception and text categorization

• Genre Identification: The parts-of-speech (POS) distribution of the review are explored by Ott et al. They used frequency count of POS tags as the features representing the review for classification.

• Detection of Psycholinguistic Deception: The psycholinguistic method approaches to assign psycholinguistic meanings to the important features of a review. Linguistic Inquiry and Word Count (LIWC) software was used by Pennebaker et al. to build their features for the reviews.

• Text Categorization: Ott et al. experimented n-gram that is now popularly used as an important feature in fake review detection. Other linguistic features are also explored. Such as, Feng et al. took lexicalized and unlexicalized syntactic features by constructing sentence parse trees for fake review detection. They show experimentally that the deep syntactic features improve the accuracy of prediction.

Sun et al. has divided these approaches into two categories.

2.1 Content Based Method:

Content based methods focus on the content of the review. That is the text of the review or what is been told in it reviews.

2.2 .Feature respected to Behaviour:

This study focuses on the reviewer that includes characteristics of the person who is giving the review. Lim et al. addressed the problem of finding users who were responsible for spam reviews. They have identified the following deceptive rating and review behaviours.

• Giving unfair rating too often: Professional spammers generally posts more fake reviews than the real ones. Suppose a product has average rating of 9.0 out of 10. But a reviewer has given 4.0 rating. Analysing the other reviews of the reviewer if we find out that he often gives this type of unfair ratings than we can detect him as a spammer.

• Giving good rating to own country's product: Sometimes people post fake reviews to promote products of own region.

For supervised classification process ground truth is determined by – helpfulness vote, rating based behaviours, using seed words, human observation etc. Sun et al. proposed a method that offers classification results through a bagging model which bags three classifiers including product word composition classifier (PWCC), TRIGRAMSSV M classifier, and BIGRAMSSV M classifier. The end result is a product word composition classifier to predict the polarity of the review. The model was used to map the words of a review into the continuous representation while concurrently integrating the product-review relations. To build the document model, they took the product word composition vectors as input and used Convolutional Neural Network CNN to build the representation model. After bagging the result with TRIGRAMSSV M classification, and BIGRAMSSV M classification they got F-Score value 0.77. However supervised method has some challenges to overcome. The following problems occur in case of supervised techniques.

• Assuring of the quality of the reviews is difficult.
• Labelled data points to train the classifier is difficult to obtain.
• Human beings are in labelling reviews as fake or genuine.

Jitendra et al. proposed semi-supervised method where labelled and unlabelled data both are trained together. They proposed to use semi-supervised method in the following situations.

• When reliable data is not available.
• Dynamic nature of online review.
• Designing heuristic rules are difficult.

The proposition includes several semi-supervised learning techniques which includes Co-training, Expectation maximization, Label Propagation and Spreading and Positive unlabelled Learning. Furthermore, the usage includes several classifiers which includes k-Nearest neighbour, Random Forest, Logistic Regression and Stochastic Gradient Descent. Using semi-supervised techniques, an accuracy of 84% has been achieved.

III. PROPOSED SYSTEM

3.1 Algorithms Used

3.1.1 Natural Language Processing (NLP):

Natural Language Processing is a field of AI in which computers understand, analyse and extract meaning from human language in a definite way. By using this developers can mould knowledge to perform tasks such as translation, extraction of relationship, analysis of sentiment, recognition of speech. It is used to Summarize blocks of text using Summarizer to extract the most important and central ideas while ignoring irrelevant information. It Uses Sentiment Analysis to identify the sentiment of a string of text, from very negative to neutral to very positive. Sentiment Analysis, can be used to identify the feeling, opinion, or belief of a statement, from very negative, to neutral, to very positive. Often, developers with use an algorithm to identify the sentiment of the term in a sentence, or use sentiment analysis to analyse social media. The great example of NLP is Social media analysis. It is also used for structuring a highly unstructured data source. Human language is astoundingly unpredictable and different. We communicate in vast manners, both verbally and recorded as a hard copy. Not exclusively are there many dialects and vernaculars, yet inside every language is a one of a kind arrangement of sentence structure and punctuation rules, terms and slang. At the point when we compose, we regularly incorrectly spell or abridge words, or preclude accentuation. At the point when we talk, we have local accents, and we murmur, falter and obtain terms from different dialects.

3.1.2 Naïve Bayes Classifier:

Naïve Bayes Classifiers are collection of algorithms based on Bayes Theorem. It is a family of algorithms which share a common principle i.e. every pair of features being classified as independent of each other. Bayes' Theorem finds the likelihood of an occasion happening given the likelihood of another occasion that has just happened. Bayes' hypothesis is expressed numerically as the accompanying condition. Figure 1 shows the pictorial representation of Naïve Bayes Classifier

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \qquad (3.1)$$

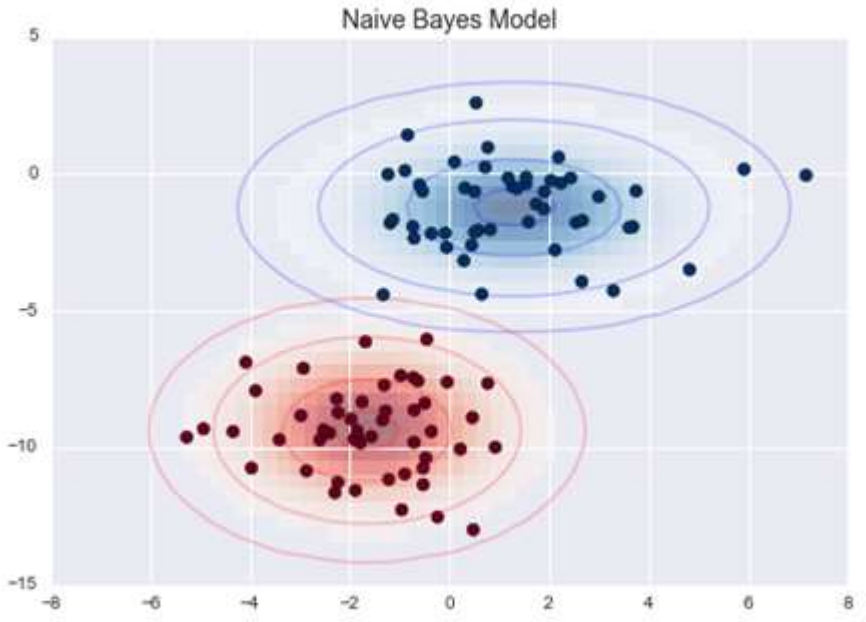

Figure 1. Naïve Bayes Classifier

IV. METHODOLOGY

For Detection of Fake online reviews we begin by using online review datasets .The dataset is split into training and testing, which is divided into the ratio 80:20 .80% of the data is used for training and 20% is used for testing the Program. Then the data is used for Natural Processing where the data undergoes three types of analysis i.e. Lexical Analysis, Syntactic Analysis, Semantic Analysis. The First type of analysis which is encountered is Lexical Analysis. Lexical Analysis is the main period of the compiler fundamentally it takes the changed source code from language pre-processors that are written as sentences. The lexical analyser breaks these syntaxes into a series of tokens, by removing any whitespaces or comments in the source code. The Second type of analysis is syntactic analysis .Syntactic analysis is also called syntax analysis which is the process of analysing a string of symbols, either in natural language, computer language or data structures, conforming to the rules of a formal grammar. The third type of analysis is Semantic Analysis. Semantic Analysis is the assignment of guaranteeing that the presentations and explanations of a program are semantically right, that is, their significance is clear and predictable with the manner by which control structures and information types should be utilized. The data which is obtained at the end of these three analysis is used for Naïve Bayes classification.

Naïve Bayes Classifiers are collection of algorithms based on Bayes Theorem. It is a family of algorithms which share a common principle i.e. every pair of features being classified as independent of each other. Naïve Bayes classifier helps to classify whether the review is fake or not. It is one of the simple and most effective Classification algorithms which helps in building the fast machine learning models that can make quick predictions. It is a probabilistic classifier, which means it predicts on the basis of the probability of an object. After training is completed the rest 20% of the dataset is used for testing purpose and that's how Naïve Bayes classifier successfully helps in classifying the review into fake or genuine. If the fake review is encountered, the program blocks the malicious user so that the fake reviews cannot be posted again. Figure 2 Represents the Flow Model of the Methodology.

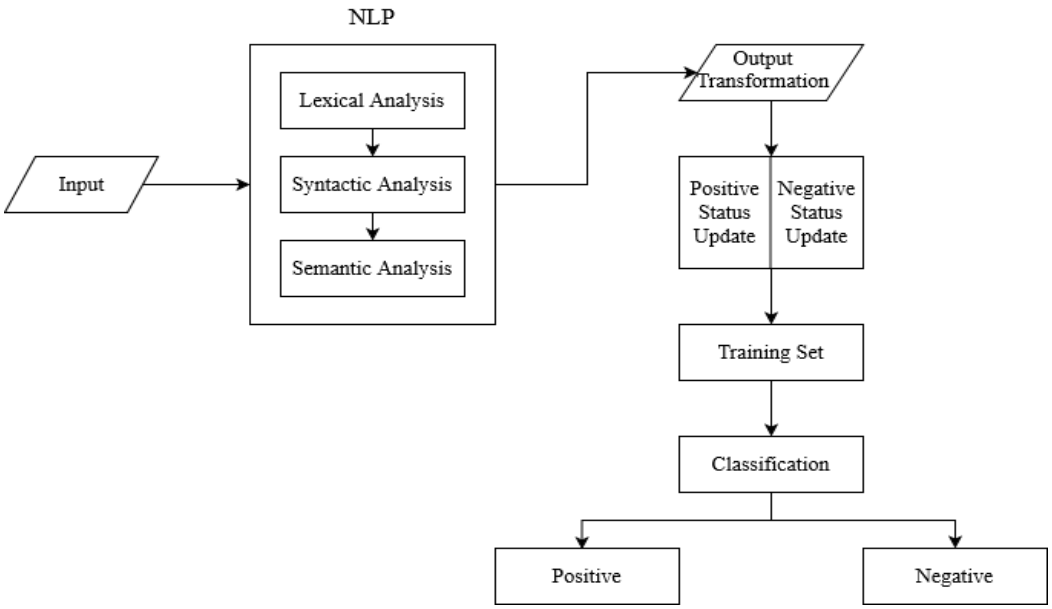

Figure 2. Proposed Methodology

Algorithm: Naïve Bayes

1. for q=1…s . u[q]=0;
2. for j=1
3. u[d[j ,p]]++;
4. for k=1
5. u[o(k-1)+(d[j, k]-1)*o(0)+d[j, p])++;
6. end for;
7. end for;

V. ADVANTAGES AND DISADVANTAGES

5.1 Naïve Bayes Algorithm

5.1.1 Advantages:

- It is easy to implement.
- Small amount of training data is required to estimate the parameters.
- Very good results is been obtained in many of the cases.
- Robust to isolate noise points.
- It handles the missing values very well.
- Robust to irrelevant attributes.

5.1.2 Disadvantages:

- Assumptions about class conditional independence, which may cause loss of accuracy of the algorithm.
- Independence assumption may not hold for some attribute.

5.2 Natural Language Processing

5.2.1 Advantages:

- The accuracy of the answer increases with the amount of relevant information provided in the question.
- The user does not need the training like to use the interface.

5.2.2 Disadvantages:

- Reliability remains an issue. It is not widely available as other forms of interface are often superior.

VI. RESULTS

Natural language processing is used for refining the dataset. We have used Naïve Bayes classifier for classifying. And dataset has been divided into a train test ratio of 80:20 for classification process. For Naïve Bayes classifier we have got 96.4% accuracy. Semi-supervised classification with EM and Positively Unlabeled learning respectively by Jiten et al, got highest accuracy of 83.00% and 83.75% for test train ratio of 80:20. Logistic regression, K-nearest neighbor, Stochastic Gradient Descent and Random Forest as classifier has been used.

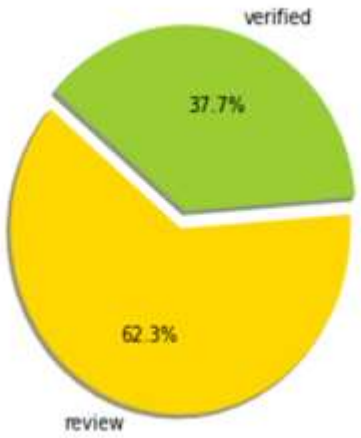

Figure 3. Verified vs. Non-Verified Reviews

Figure 3 shows a pie chart plotted for the review column and verified column. The verified column is 37.7% and review column is 62.3%. And Figure 4 shows a bar graph plotted for label, product category against number of occurrences. It shows how many products has occurred how many times.

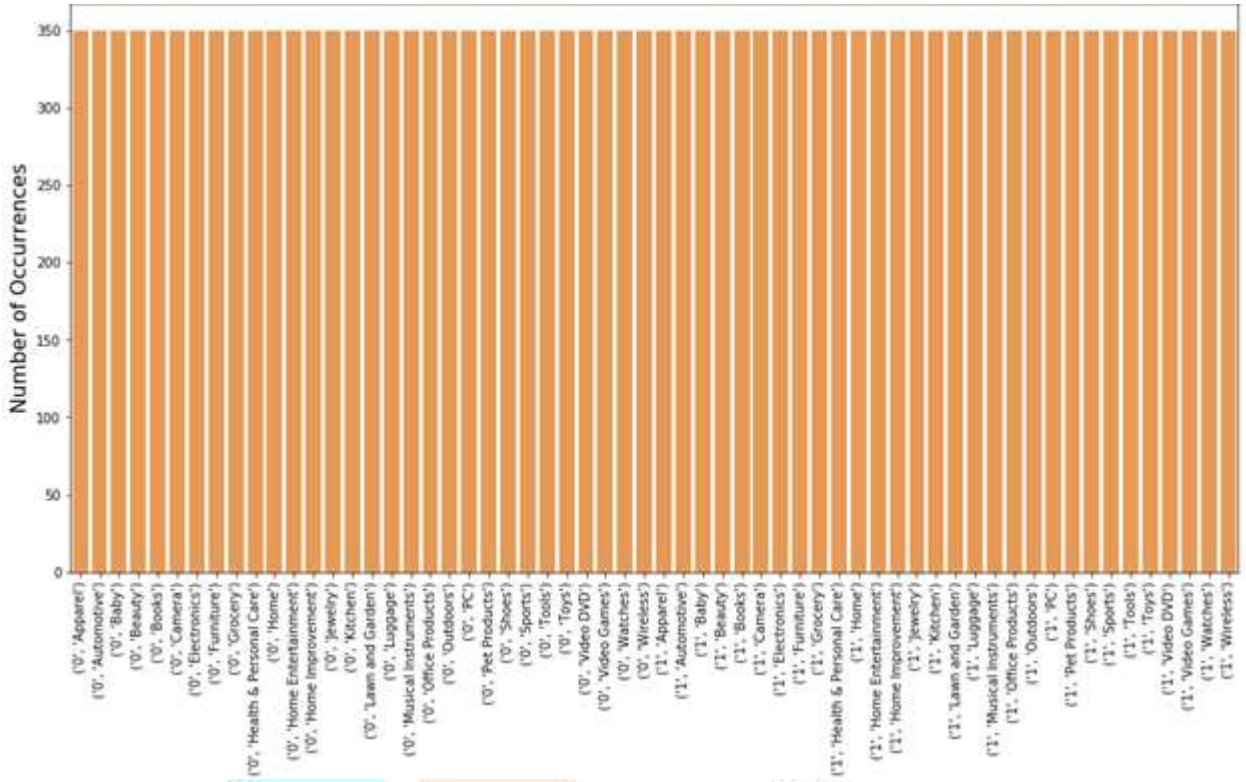

Figure 4. Label vs Product Category

VII. CONCLUSION

In these days, determining whether a review is true or false has been one of the greatest challenges, such reviews can play a major role in purchasing decision of a customer. As a result, a lot of work has been done on the same. We have emphasized different methods in this paper to distinguish fake reviews from a collection of online reviews. The fundamental aim of this paper is to improve the accuracy and efficiency. We have considered the results about the same from previous studies and have come up with an optimal solution. The source data of reviews are from Amazon marketplace. We have made use of linguistic approach such as content of the review and the length of the review as parameters. The conclusion is that of all the classifying algorithms, Naïve-Bayes is more promising.

VIII. FUTURE WORK

As far now, we have limited our research focusing more on the content of the reviews. Future work involves not only reviews, but also behavior of the person based on his other reviews on the same platform, as a result we can come up with a better understanding of an individual behavioral pattern which can further increase the accuracy of determination and segregation of reviews.

REFERENCES

[1] Chengai Sun, Qiaolin Du and Gang Tian, "Exploiting Product Related Review Features for Fake Review Detection" Mathematical Problems in Engineering, 2016.

[2] M. Ott, Y. Choi, D. Cardie, and J. I. Hanckit, "Finding Deceptive Opinion Spam by Any Stretch of the Imagination" in Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics.

[3] J. W. Pennebaker, M. E. Francis, and R. J. Booth,"Linguistic Inquiry and Word Count: Liwc" vol. 71, 2001.

[4] S. Feng, R. Banerjee, and L. Choi, "Syntactic stylometry for deception detection".

[5] J. Li, M. Ott, C. Cardie, and H. Honyic, "Towards a General Rule for Identifying Deceptive Opinion Spam" in Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (ACL), 2014.

[6] E. P. Lim, V.-A. Nguyen, N. Jindal, B. Liu, and H. W. Lawnic, "Detecting Product Review Spammers Using Rating Behaviors" in Proceedings of the 19th ACM International Conference on Information and Knowledge Management (CIKM), 2010.

[7] J. K. Rout, A. Dalmiana, and K.-K. H. Choo.D, "Revisiting Semi-Supervised Learning for Online Deceptive Review Detection" IEEE Access, Vol. 5, pp. 1319–1327, 2017.

[8] J. Karimpour, A. B. Noroozi, and S. Alizey, "Web Spam Detection by Learning from Small Labeled Samples" International Journal of Computer Applications, vol. 50, no. 21, pp. 1–5, July 2012.