FAKE REVIEW MONITORING AND AUTOMATED CHATBOT

Muskan Goyal, Neha Gharat, Siddhesh Kohale, Pratik Beloshe

Prof. Krishnendu Nair ,Department of Computer, PCE, Navi Mumbai, India - 410206

Abstract - E-commerce is the online buying and selling process that is extremely important in our daily life now. One of the important aspects of E-commerce is Reviews. Most of the time, customers place the order only by reading the reviews related to the particular product. In this situation, it might be possible that the reviews are fake, because of which the user might get second thoughts of not buying the product. So, to remove this type of fake reviews and provide the users with the original reviews and ratings related to the products, We proposed a Fake Product Review Monitoring and Removal System by using different approaches including identification process, opinion mining, and machine-learning approach. The website will have also have a chatbot system that provides an automatic response to the incoming customer to-seller question. In general, the proposed. The system consists of two main agents: communication and the intelligent part. Since we rely greatly on online shopping, nowadays, providing customers with the best experience is the priority. The Chatbot will reply to the Customers product-related Query. The objective of this project is to provide users with the best experience of products and websites by removing fake reviews and having an automatic chatbot to easily solve their Queries.

Key Words: **Fake Reviews, chatbot.**

1.INTRODUCTION

There are different ways to shop like you can buy a specific thing of your need by going to a store or mall. In this style of shopping, the seller gives you the feedback of the product, you do not know whether he/she is giving fake feedback or original. Because, it is upon seller honesty, how much the seller is true in his/her words and you have to carefully examine the product because you do not have any other option in examining the product. If you don't pay attention in buying that product then it may be proved a waste for you. On the other hand, nowadays the source of shopping has been changed. You can buy the products from the online stores of different brands. Here, you have to place the order without seeing and examining the original product. You read the reviews and buy the product. Therefore, you are dependent on the reviews about the product. These reviews may be original or fake. The customer wants to buy an original and reliable product, which is possible only when you get the original feedback related to that product.

1.1 Literature Survey

A. Fake Review Monitoring using opinion mining(2018):
There have been several attempts for spam review detection till today. In this paper, we propose a general framework to detect spam reviews. Sentiment analysis or opinion mining is a field of study that analyzes people's sentiments, feelings, or emotions towards certain entities. The software will do analysis and then if any fake review is found from any IP address consistently then the admin user can block that IP address. It also sends mail. And user can be sure about the product's availability on that application and reviews too.

B. Fake News Detection using Naïve Bayes Classifier(2017):
The research showed that even quite a simple Artificial Intelligence algorithm may show a good result on such an important problem as fake news classification. Therefore the results of this research suggest even more, that artificial intelligence techniques may be successfully used to tackle this important problem.

C. Spam Review Detection Using the Linguistic and Spammer Behavioural Methods (2020): The applications of the study include spam review detection in product/services reviews on e-commerce websites product/services websites e.g. Amazon, etc. Furthermore, additional attributes will be added to the dataset to improve the accuracy and reliability of the spam review detection models. These may include an IP address of the spammer, registered email address, and signed-in location.
D. Fake news detection on social media(2017): In this article, we explored the fake news problem by reviewing existing literature in two phases: characterization and detection. In the characterization phase, we introduced the basic concepts and

principles of fake news in both traditional media and social media. In the detection phase, we reviewed existing fake news detection approaches from a data mining perspective, including feature extraction and model construction.

E. Intelligent Interface for Fake Product Review and Removal (2019): In the proposed work, the dataset is developed that contains Urdu and Roman Urdu reviews. It is difficult to detect fake reviews by yourself. So, the ngram approach is used to detect fake reviews for multiple languages. It is observed that the text categorization with SVM classifier s the best approach for the detection of fake reviews.

F. Fake Review Detection using Machine Learning Techniques (2018): Due to the rapid development of the internet, the size of the reviews of the items/products increases. These huge amounts of information are generated on the Internet; there is no analysis of the quality of reviews that are written by the consumer. Anyone can write anything which conclusively leads to fake reviews or some companies are hiring people t post reviews. Some of the fake reviews that have been intentionally fabricated seem genuine, the capability to identify fake online reviews is crucial. In this paper, we have discussed different fake review detection techniques that are based on unsupervised, supervised as well as semi-supervised methodologies. In this paper, we have seen different features in detail like linguistic features, behavioral and relational features. We have also compared different techniques to identify fake reviews. We have also discussed the major challenges of fake review detection.

1.2 Summary of Related Work

There are many projects and research papers published on this topic that involves fake review monitoring, but all of them have either used opinion mining, IP tracking, etc. Here we have listed some of the papers below with their datasets and success ratio.

Advantages:

- The user gets genuine reviews about the product.
- User can post their own review about the product.
- Users can spend money on valuable products.

Disadvantages:

- If the social media optimization team uses the different IP addresses to send the review, the system will fail to track the fake review.

2. System Architecture

The system architecture is given in Figure 1.

Each block is described in this Section.

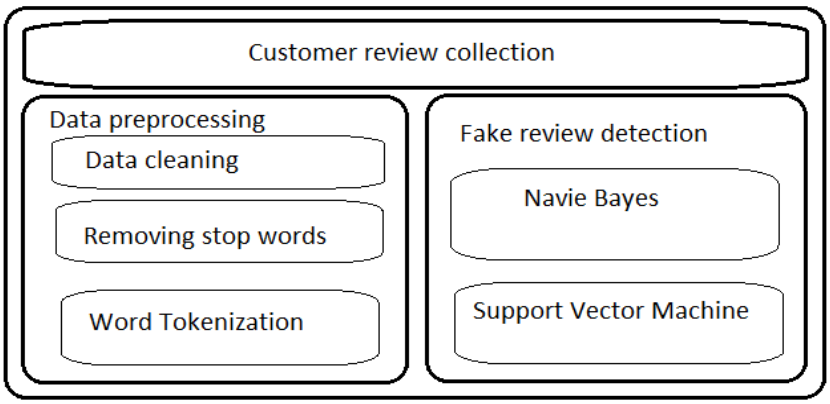

Fig. 1 Proposed system architecture for fake reviews

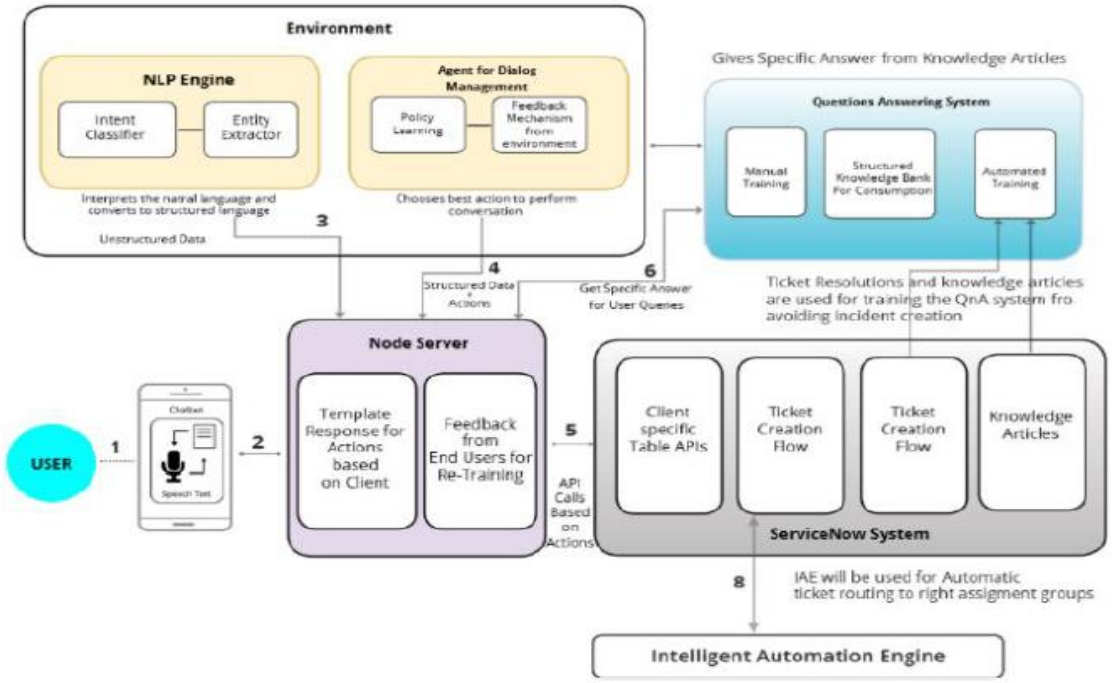

Fig 2 Proposed system for automated Chatbot

A. Customer Review Collection: The first part is Consumer review data collection in which raw review data was collected from sources, such as Amazon, Flipkart, etc. reviews. Doing so was to increase the diversity of the review data.

B. Data Pre processing: Processing and refining the data by removal of irrelevant and redundant information as well as noisy and unreliable data from the review dataset.

- Step 1 Sentence tokenization

The entire review is given as input and it is tokenized into sentences using the NLTK package.

- Step 2 Removal of punctuation marks

Punctuation marks used at the starting and ending of the reviews are removed along with additional white spaces.

- Step 3 Word Tokenization

Each individual review is tokenized into words and stored in a list for easier retrieval.

- Step 4 Removal of stop words

Affixes are removed from the stem. For example, the stem of "cooking" is "cook", and the stemming algorithm knows that the "ing" *suffix* can be removed.

C. Fake Review Detection: After cleaning the data and getting the vectors for the same, we will split the data generally in a ratio of 80:20. So, 80% of the data will be used for training the model and the remaining 20% will be used for validation of the model and eventually getting its accuracy. Now, we will be training our model with the training data which we have prepared. Since this problem is a text classification problem, we will be using the naive Bayes algorithm as it works best for text classific ation problems. After the model is trained, we will predict the target variable for the test set and compare it with the actual data to evaluate the model and find its accuracy.

D. Naive Bayes Algorithm: Naive Bayes is a simple technique for constructing classifiers: models that assign class labels to problem instances, represented as vectors of feature values, where the class labels are drawn from some finite set[14]. Abstractly Naive Bayes is a conditional probabilistic model: given a problem instance to be classified, represented by a vector x=(x1,x2,. xn) representing some *n* features (independent variables), it assigns to this

$p(C_k \mid x_1,\ldots,x_n)$ $p(C_k \mid x_1,\ldots,x_n)$

Instance probabilities

For each K possible outcomes or classes *Ck*.

The problem with the above formulation is that if the number of features *n* is large or if a feature can take on a large number of values, then basing such a model on probability tables is infeasible. We, therefore, reformulate the model to make it more tractable

E. Support Vector Machine: In machine learning, support vector machines[15] are supervised learning models with associated learning algorithms that analyze data used for classification and regression analysis. Given a set of training examples, each marked as belonging to one or the other of two categories, an SVM training algorithm builds a model that assigns new examples to one category or the other, making it a non-probabilistic binary linear classifier (although methods such as Platt scaling exist to use SVM in a probabilistic classification setting). An SVM model is a representation of the examples as points in space, mapped so that the examples of the separate categories are divided by a clear gap that is as wide as possible.

F. Output Block Description: After the fake reviews detected, we need to remove them so that users can spend his/her money on a quality product.

3. CONCLUSIONS

With the increasing popularity of online shopping, more and more people do online shopping instead of traditional purchase. However, the E-commerce domain has also been used to spread fake reviews, which has strong negative impacts on individual users and broader society on buying the product.
Our website will do analysis and then post genuine reviews on the product. And users can be sure about the products availability on that website and reviews too. It will help the user to pay for the right product without any getting into scams. In future we would try to improve the method of calculating the sentiment score of the reviews. We will also update the website with new technologies. We also further discussed the datasets, evaluation metrics, and promising future directions in fake reviews detection research and expand the field to other applications

ACKNOWLEDGEMENT

It is our privilege to express our sincerest regards to our supervisor Prof. Krishnendu Nair for the valuable inputs, guidance, encouragement, whole-hearted cooperation and constructive criticism throughout the duration of this work. We deeply express our sincere thanks our Head of the Department Dr. Sharvari Govilkar and our Principal
Dr. Sandeep M. Joshi for encouraging and allowing us to presenting this work.

REFERENCES

1. https://www.sciencedirect.com/science/article/pii/S030645731730657X
2. https://ieeexplore.ieee.org/abstract/document/6483152/
3. https://ieeexplore.ieee.org/document/8679186.
4. https://ieeexplore.ieee.org/document/8592630