International Journal of Engineering Applied Sciences and Technology, 2022
Vol. 7, Issue 12, ISSN No. 2455-2143, Pages 47-52
Published Online April 2023 in IJEAST (http://www.ijeast.com)

MONITORING AND REMOVAL OF FAKE PRODUCT REVIEW USING MACHINE LEARNING (ML)

Swarnajit Bhattacharya
[*1]Department of Electronics and Instrumentation (AEIE),
Haldia Institute of Technology, Haldia

Abstract: **The employment of Natural Language Processing (NLP) as a means for identifying and eliminating counterfeit products from a dataset is a significant endeavor. The current project employs a Machine Learning (ML) Algorithm to discern fraudulent customer reviews within a given dataset. The algorithm leverages predictive modeling techniques to estimate the authenticity of such reviews, and subsequently assess their validity. The prominence of product reviews as a critical determinant of consumer purchasing behavior in the realm of electronic commerce has led to a concerning escalation in the frequency of counterfeit reviews on websites and applications. It is imperative for the prominent E-commerce firms to confront the issue of counterfeit product reviews, prior to engaging in the procurement of commodities from established entities. The utilization of this particular approach facilitates the rectification process concerning counterfeit product evaluations, thereby eliminating the prevalence of spammers. This proactive measure works to ensure that there is no potential compromise of trust among users of E-commerce platforms. Through implementation of this initiative, the enterprise's administration can identify spurious evaluations and subsequently implement appropriate measures directed towards their resolution. The present model has been constructed employing the Naïve Bayes Algorithmic approach. By implementing the algorithm, it is possible to discern spam reviews from those that are legitimate, on websites or applications. To tally the individuals who engage in fraudulent online practices, a dataset is necessary. In our study, the "Amazon Academic Dataset" serves as the fitting reference point to prepare the model and can potentially increase the precision and adaptability of the results.**

I. INTRODUCTION

In contemporary times, online shopping has witnessed substantial growth in popularity, with people increasingly relying on this platform to purchase a diverse range of goods and services. In the wake of the COVID-19 pandemic and resultant lockdowns, an increasingly significant proportion of individuals have come to rely exclusively on online markets. The proliferation of e-commerce has led to the emergence of numerous new enterprises. The emergence of newly established startup companies is often accompanied by the pernicious presence of spammers who seek to defraud unsuspecting individuals.

When making a purchase, individuals in the general population typically do not engage in a physical inspection of the product to assess its quality. Instead, a pervasive reliance on online reviews and ratings bestowed by fellow consumers has become the norm. In contemporary online commerce, the general public primarily assesses a product's quality through its ratings and the number of individuals who have evaluated it, which are closely scrutinized at the moment. Alternatively, in certain instances, individuals may simply assess the quantity of individuals providing feedback regarding a specific product. The weather was very bad yesterday. It was raining heavily and the wind was blowing hard. Many trees were knocked down and there were floods on the streets. Rewritten: The adverse weather conditions experienced yesterday were characterized by heavy rainfall and strong winds that resulted in the uprooting of several trees and flooding of the streets.

There exists a prevalent practice wherein spammers engage in the usage of unsolicited electronic communication, commonly referred to as spamming, as a means of improving their respective company products. The proliferation of deleterious comments on competing company products by individuals engaging in spamming activities has emerged as a paramount challenge for enterprises that exclusively rely on digital marketing strategies.

The act of spammers providing fraudulent reviews, featuring laudatory adjectives such as "awesome," "fantastic," or "very good," is prevalent in modern society. The aforementioned terms denote the prime lexicon employed by purveyors of unsolicited messages to soliciting increased commentary on merchandise. This matter has become of significant concern. It is imperative to employ effective methods that can identify individuals who engage in spamming activities of this nature. Amazon and Flipkart, as commercial entities, possess substantial inventories of products, vast data sets, and significant volumes of customer feedback.

The manual detection and removal of spammers in a sizeable dataset is deemed infeasible. The enhancement of technology and model training can be achieved through the utilization of Machine Learning algorithms. Additionally, Natural Language Tool Kit (NLTK) can aid in data filtration, and the Naïve Bayes algorithm can be employed to accomplish this. Spamming comments can be accurately detected and subsequently eliminated from the dataset.

The Natural Language Toolkit (NLTK) comprises a predetermined set of words intended for use in the training of a model. These NLTK packages are imported and applied in algorithm implementation, wherein sentiment analysis is performed on the model. The outcome of this process determines whether a review is classified as positive or negative.

Sentiment analysis involves subjecting each individual word to an algorithm prior to its implementation. Prior to the sentiment analysis, pre-processed data that has had stopwords extracted through the use of the NLTK packages is utilized. The sentiment analysis is exclusively conducted on the value-added words for analysis. Subsequent to obtaining the outcomes, it is imperative to determine the efficacy, as it provides insight into the performance of the model. The present investigation endeavors to identify spammers in the dataset. In order to achieve this objective, a spamming test was conducted subsequent to the model's classification. As a result, the reviews authored by spammers were successfully identified in the dataset.

Subsequently, fabricated reviews are systematically eliminated from the dataset. To train the model, the "Amazon academic dataset" is being utilized - a dataset comprising both inauthentic and genuine reviews. The significance of this specific dataset lies in its crucial role in facilitating the training of the model with the algorithm.

II. LITERATURE SURVEY

2.1. Fake review detection through supervised classification.
1.Pankaj Chaudhary, 2.Abhimanyu Tyagi 3.Santosh

Over the course of several years, research has been conducted on the means by which information contained within reviews is tackled via social media. The veracity and precision of the reviews, as well as the contextual framework and various other amalgamations of multiple dimensions, have not been observed. The compilation of reviews encompasses multiple dimensions, which entails sourcing them from various channels to evaluate their efficacy, precision, and veracity. Primarily, the data is acquired from social media platforms. It is widely recognized that social media platforms, such as websites, are commonly utilized and serve as a primary venue for individuals to express their opinions and evaluations. The utilization of their methods by spammers to spam reviews has the potential to be detrimental to the online health landscape. The present study primarily sourced its data utilizing a data approach technique that involved the classification of credibility as evaluated by employees. Notably, the supervised classification method was employed to categorize reviews, with training the model using algorithms to detect and classify spam within such reviews. The given text has already been written in an academic way of writing.

2.2. Opinion spam and analysis.

Nitin Jindal and Bing Liu are affiliated with the Department of Computer Science at the University of Illinois at Chicago.

This analysis is founded on users' expressed opinions regarding various products. At present, the analysis involves an examination of the sentiment of the language in the reviews, thereby determining whether they convey a positive or negative slant. Additionally, the identification of opinion spam is possible via the utilization of an opinion analysis technique. Numerous individuals utilize websites to engage in various online activities, however, some resort to performing web spam tactics. The state of the economy is in a precarious condition due to the ongoing pandemic, with unemployment rates at an all-time high and businesses struggling to stay afloat. The government has implemented various measures to mitigate the impact, such as issuing stimulus checks and providing loans to small businesses. However, these efforts may not be enough to fully alleviate the economic woes faced by the country.

The primary objective of web spammers is to direct their attention towards websites and generate spam content that appears to be either a positive endorsement or a negative criticism. In order to impede such spam activities, a process of sentimental analysis is performed on reviews to detect spammers in a timely manner and subsequently execute measures such as deactivating or permanently blocking their social media accounts.

III. METHODOLOGY

The model has been designed to effectively identify reviews that lack authenticity and subsequently eliminate such reviews from the review list. It is imperative that the reviews undertake an analysis of the acquired data. Initially, the creation of the website was aimed at acquiring website data. The administrative task at hand involves incorporating categories into the system. The administrator holds the ability to incorporate an infinite number of categories, along with their corresponding details.

Subsequently, the products are allocated to their respective category, in which each product is identified by its name, description, rate, quantity, and image. Subsequently, on the succeeding page, the administrator is able to preview the visual appearance of the merchandise. Subsequently, the number of registered customers utilizing the aforementioned website can be cross-referenced by the observer.

Subsequently, customer reviews can be obtained and exhibited, thereby enabling the administrator to access comprehensive information pertaining to the customer's feedback on the product. The administrator has the ability to log out from the current page, thereby redirecting the user to the homepage.

Thereafter, the role of the customer shall ensue. It is imperative for patrons to complete the registration process in order to gain access to the website. On the registration page, the customer is required to provide their full name, email address and phone number, as well as establish their own unique login password. The following is an example of a text rewritten in an academic style: Original Text: Yo, I went to the store to grab some chips and ended up running into my ex-girlfriend. She was with some dude I didn't know, and I could tell she was trying to make me jealous. It was super annoying. Rewritten Text: During a recent visit to the local convenience store, I found myself unexpectedly encountering a former romantic partner. The individual in question was accompanied by an unfamiliar companion, and I sensed an apparent attempt on her part to provoke feelings of envy within me. This circumstance was highly vexing.

These particulars can be observed by the administrator. Subsequent to registering in the online platform, customers will be able to access the login section featuring a range of products authorized by the site administrator. Customers have the option to select desired products and also have the capacity to augment the quantity of products they intend to purchase to the shopping cart, following which they are able to conclude the process by submitting the page.

The rate is computed and presented on the webpage for the client's purchase of the product. Upon purchasing the respective products, customers are afforded the opportunity to append reviews to said products. The reviews that have been provided are subjected to subsequent analysis to ascertain their authenticity.

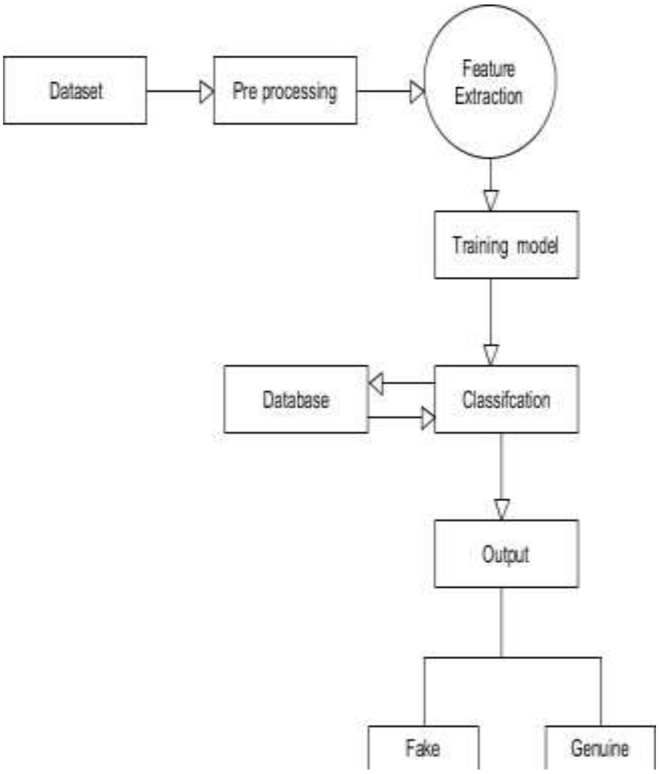


Fig 1: Flow chart diagram of Monitoring and removal of fake product

In this particular phase of the market flow, a substantial number of consumers engage in the practice of posting spam reviews with the aim of disparaging the products. Alternatively, favorable feedback may be provided for substandard products. The spammers habitually create counterfeit evaluations in order to propagate adverse appraisals on web-based platforms. The reviews furnished by patrons are subjected to sentiment analysis methodology and assessed by the administrator through the Naive Bayes algorithm to calculate the precision of the reviews following model training.

Subsequent to this stage, critical evaluations shall be subjected to analysis with a view to identifying the application of "stopwords" that have been incorporated. These "stopwords" are derived from the NLTK (Natural Language Tool Kit), a sizable database of approximately 50 megabytes that has been installed to facilitate the said process. The aforementioned are examples of English lexical items such as "The," "is," and "was." The aforementioned lexicons hold nominal import. Upon installation and subsequent comparison with our dataset, this method effectively eliminates the stopwords contained therein. This procedural step is commonly referred to as the feature extraction process. The elimination of punctuation, specifically full stops and commas, is referred to as the pre-processing technique. The data cleaning process entails two distinct steps.

Subsequently, once the data is obtained, it is amenable for model training. The process of data training is bifurcated into two fundamental components, wherein a portion is deemed suitable for the actual training of the data, while the other section is reserved for the purpose of data testing. The model's training and testing are conducted in a ratio of 80:20 percent. In order to conduct sentiment analysis, the requisite tool accessible in Python is being imported. The analysis of the featured data will be conducted through the utilization of sentiment analysis. The present study undertakes an analysis of individual words and assigns random numerical values to each word.

The Naïve Bayes algorithm is employed to train the model. In comparison with alternative algorithms, the Naïve Bayes algorithm has been found to be the most effective means of text data classification. The mathematical procedure is founded on the likelihood of affirmative and opposing information.

P (a / b) = (P (b / a) * P (a)) / P (b)

The aforesaid arrangement is the methodology utilized to categorize the outcome as either a constructive appraisal or a detrimental critique. Upon examining any given sentence, it must first undergo a process of pre-processing in order to filter out extraneous information. Only those words that have passed through the filtration stage are deemed useful for subsequent algorithmic implementation.

Sentence[x1, x2, x3,………xn]…….(1)

The x prepresents the words present in the sentence, the sentence is divided by its words for further processing. then we use probability, divided by the number of words in the sentence.

P (y = yes / sentence)…….(2)

Implementing (1) in (2) we get,

= P (y = yes / x1, x2, x3……xn) α P (y) * $\prod n$

= P (x1 / y = yes) * P (x2 / y = yes) *………….. P(xn / y = yes)

= P (y = yes) * P (x1 / y = yes)

In the context of probability theory, the computation of the probability of a given event may involve dividing the cardinality of the subset consisting of positive outcomes by the total number of data points in the sample space. Similarly, the probability of the complementary event may be obtained by dividing the cardinality of the subset of negative outcomes by the total number of data points in the sample space. Conducting probability calculations enables the derivation of precise solutions.

The numerical value generated through a probability-based process is presented as follows:

- For the positive word, gives random number between 4 to 5.
- For the neutral words, gives random number between 2 to 3.
- For the negative words, gives random number between 0 to 1.

Using the randint() method, the random number is assigned to the single words in sentimental analysis.

The dataset comprises an assortment of diverse data. Within the realm of data collection, there exist primarily three distinct categories of dataset accumulation, namely:

1) taking the data from the websites,
2) taking data from manually,
3) inbuilt taking of data. Here we are taking the data from the website. Which is a collection of fake and non fake mixture of data.

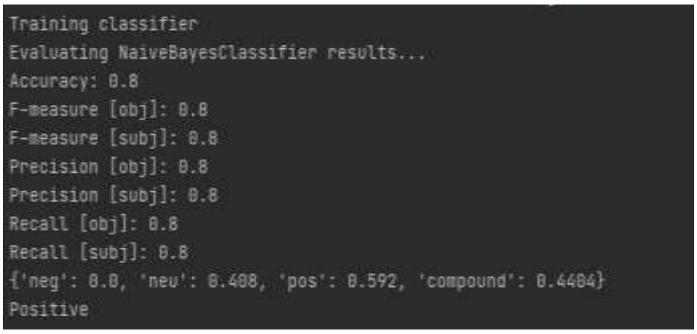
The aforementioned repository comprises customer feedback on the product at hand. The dataset employed in this study comprises approximately 88,000 reviews, sourced from both Amazon's website and manual curation. This dataset includes several associated features, such as timestamps, voting records, and identification codes. Specifically, this dataset is utilized to perform the detection and removal of fraudulent product reviews.

The removal system takes a given dataset and initiates a process of dataset mining followed by the imposition of a 1800 millisecond time-stamp. During this time-frame, any review that is found to be repeated six or more times is qualified as a spam or fake review and consequently eliminated from the dataset. The present study outlines the methodology employed to identify and eliminate reviews from the dataset.

IV. RESULTS AND DISCUSSION

As previously noted, the methodology has been implemented to address the presence of fraudulent product reviews in the dataset. By employing the Naïve Bayes algorithm, optimal values pertaining to accuracy, F-measure, precision, and recall can be obtained. The precision of the system stands at approximately 80%. As previously discussed, the provision of examples serves to identify whether a given situation is indicative of a positive or negative outcome. Prior to evaluating the quality of a shirt, preprocessing is necessary. Upon conducting such preliminary measures, it can be deduced that the shirt exhibits exceptional quality. The punctuation marks have been extracted from the sentence. Subsequently, the process of feature extraction. The sentence was subjected to a comparison with stopwords, following which the words "is" and "of" were subsequently eliminated from the sentence. The aforementioned approach adheres to the conventions of academic writing by utilizing a more formal and structured tone.

Subsequently, the process of sentiment analysis shall be executed, wherein the categorization "shirt" receives a neutral value. The term "very" denotes a neutral value while "good" is a representation of a positive value. The concept of "quality" is considered a value that is essentially impartial. Subsequently, the Naïve Bayes algorithm shall be employed for its application. This algorithm generates stochastic values for the lexicon and computes the corresponding probability distribution function. This shall yield the ultimate outcome pertaining to the aforementioned matter.

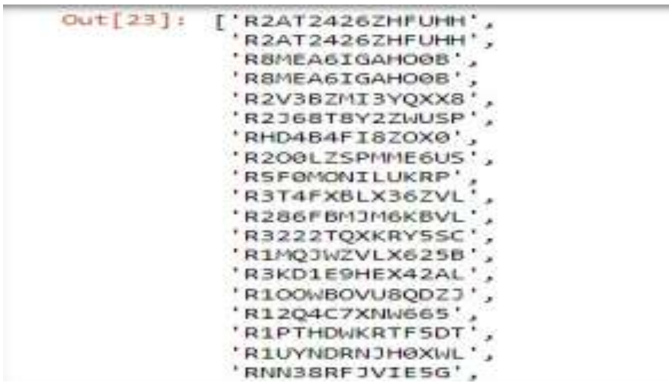


```
Training classifier
Evaluating NaiveBayesClassifier results...
Accuracy: 0.8
F-measure [obj]: 0.8
F-measure [subj]: 0.8
Precision [obj]: 0.8
Precision [subj]: 0.8
Recall [obj]: 0.8
Recall [subj]: 0.8
{'neg': 0.0, 'neu': 0.408, 'pos': 0.592, 'compound': 0.4404}
Positive
```

Fig 2: Result showing accuracy using Naïve bayes algorithm

The image above presents the values for Accuracy, F-measure, precision, and recall, alongside values specific to neutral and positive lexical items, and culminates in the calculation of a compound value. The observed metric for accuracy approximates 0.8.

In order to eradicate inauthentic reviews from the dataset, the present study entails analyzing the reviews submitted by users on the website. Such reviews are subject to pre-processing measures, such as feature extraction, with the subsequent implementation of imported NLTK packages. The ultimate objective is to eliminate fraudulent reviews from the dataset. Residential education can have a significant impact on the development of individuals, particularly in regards to their personal growth, academic achievement, and social skills. Being surrounded by a community of peers and engaging in academic, extracurricular, and social activities can enhance one's educational experience and broaden their perspective, ultimately resulting in a more well-rounded and well-prepared individual.



```
Out[23]:  ['R2AT2426ZHFUHH',
           'R2AT2426ZHFUHH',
           'R8MEA6IGAHO0B',
           'R8MEA6IGAHO0B',
           'R2V3BZMI3YQXX8',
           'R2J68T8Y2ZWUSP',
           'RHD4B4FI8ZOX0',
           'R2O0LZSPMME6US',
           'R5F0MONILUKRP',
           'R3T4FXBLX36ZVL',
           'R286FBMJM6K8VL',
           'R3222TQXKRY5SC',
           'R1MQJWZVLX625B',
           'R3KD1E9HEX42AL',
           'R1OOWBOVU8QDZJ',
           'R12Q4C7XNW665',
           'R1PTHDWKRTF5DT',
           'R1UYNDRNJH0XWL',
           'RNN38RFJVIE5G',
```

Fig 3: Result showing the removed reviews from the dataset

The image presented above displays the reviews that have been excluded from the dataset due to their repetitive nature. These reviews qualify as spam, as they are repeated with a frequency of at least six occurrences within a given minute. The outcomes are exhibited within a solitary vertical arrangement.

V. CONCLUSION

As the contemporary landscape progresses, there is a steady and exponential increase in the amount of data published on websites. The advent of social media has led to the prolific generation of copious amounts of data in the form of reviews, comments, customer feedback, and star ratings on a perpetual basis. The vast quantity of data generated by users is rendered inconsequential without the implementation of data mining methodologies. Given the pervasive prevalence of fraudulent reviews, the implementation of the present application effectively facilitates the identification and elimination of spurious reviews in a highly efficient manner. The issue of spotting fraudulent product reviews has been properly examined, providing a judicious understanding of its legality and necessity for both administrators and e-commerce entities. The objective is to identify a Naïve Bayes algorithm that is suitable for accomplishing the task of counterfeit product review detection and its consequent removal.

VI. ACKNOWLEDGEMNT

I extend my gratitude to Mrinmoy Sen, an esteemed faculty member specializing in Computer Science Engineering (CSE) at the prestigious Haldia Institute of Technology, for his invaluable guidance and support in my research endeavors. Furthermore, I am particularly indebted to my Mathematics instructor, Mrs. Twisha Das, for her unwavering assistance and contributions to this area of study. She played a pivotal role in fortifying my mathematical competencies by imparting diligent teaching and aiding in the comprehension of complex algorithms. Without her unwavering dedication and diligence, the successful creation of the mathematical model designed for the detection of fraudulent product reviews would not have been achievable. Finally, my sincere gratitude is extended to my parents who continually remained steadfast in extending their support to me throughout the course of this research article. I underscore that the success of this research work would not have been achievable without their unwavering contributions.

VII. REFERENCES

[1] "Fake review detection using opinion mining" by Dhairya Patel, Aishwerya Kapoor and Sameet Sonawane, International Research journal of Engineering and technology (IRJET), volume 5, issue 12, Dec 2018.

[2] McCallum, Andrew. "Graphical Models, Lecture2: Bayesian Network Represention" (PDF). Retrieved 22 October 2019.

[3] "Fake review detection from product review using modified method of iterative computation framework", by EkaDyar Wahyuni & Arif Djunaidy, MATEC web conferences 58.03003(2016) BISSTECH 2015.

[4] Rajashree S. Jadhav, Prof. Deipali V. Gore, "A New Approach for Identifying Manipulated Online Reviews using Decision Tree ". (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 5(2), pp 1447-1450, 2014

[5] Long- Sheng Chen, Jui-Yu Lin, "A study on Review Manipulation Classification using Decision Tree", Kuala Lumpur, Malaysia, pp 3-5, IEEE conference publication, 2013.

[6] Benjamin Snyder and Regina Brazil, "Multiple Aspect ranking using the Good Grief Algorithm "Computer Science and Artificial Intelligence Laboratory Massachusetts Institute of Technology2007.

[7] Ivan Tetovo, "A Joint Model of Text and Aspect Ratings for Sentiment Summarization "Ivan Department of Computer Science University of Illinois at Urbana, 2011.

[8] N. Jindal and B. Liu, "Analyzing and detecting review spam," International Conference on Web Search and Data Mining, 2007, pp. 547- 552.

[9] N. Jindal and B. Liu, "Opinion spam and analysis," International Conference on Web Search and Data Mining, 2008, pp. 219-230 .