VISVESVARAYA TECHNOLOGICAL UNIVERSITY
JNANA SANGAMA, BELGAUM – 590014

A Seminar Report on

OPINION MINING BASED FAKE PRODUCT REVIEW MONITORING AND REMOVAL SYSTEM

Submitted in partial fulfillment of the requirements for the award of degree of

BACHELOR OF ENGINEERING IN
COMPUTER SCIENCE AND ENGINEERING

by
SHIVATHMAJ SHENOY M
4CB18CS085

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
CANARA ENGINEERING COLLEGE
BANTWAL 574 219
2021-22

Certificate

Certified that the seminar report entitled **Opinion Mining based Fake Product review Monitoring and Removal System** is a work carried out by **Shivathmaj Shenoy M (4CB18CS085),** in partial fulfillment of the requirements for the award of the degree of **Bachelor of Engineering (BE) in Computer Science and Engineering (CSE) of Visvesvaraya Technological University, Belagavi**, during the year 2021-2022. Further it is certified that all the corrections/suggestions indicated during the internal assessment have been incorporated in this report. The report has been approved and accepted as it satisfies the academic requirements (partial) in respect of seminar (18CSS84) prescribed for the aforesaid degree.

_____
Signature of the
Coordinator

_____
Signature of the
HOD

ABSTRACT

Fake review detection and its elimination from the given dataset using different Natural Language Processing (NLP) techniques are important in several aspects. In this article, the fake review dataset is trained by applying two different Machine Learning (ML) models to predict the accuracy of how genuine is the reviews in a given dataset. The rate of fake reviews in Ecommerce industry and even other platforms is increasing when depend on product reviews for the item found online on different websites and applications. The products of the company were trusted before making a purchase. So this fake review problem must be addressed so that these large E-commerce industries such as Flipkart, Amazon, etc. can rectify this issue so that the fake reviewers and spammers are eliminated to prevent users from losing trust on online shopping platforms. This model can be used by websites and applications with few thousands of users where it can predict the authenticity of the review based on which the website owners can take necessary action towards them. This model is developed using Naïve Bayes and random forest methods. By applying these models one can know the number of spam reviews on a website or application instantly. To counter such spammers, a sophisticated model is required in which a need to be trained on millions of reviews. In this work "amazon Yelp dataset" is used to train the models and its very small dataset is used for training on a very small scale and can be scaled to get high accuracy and flexibility.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Now a day's people have started to give fake reviews on the products. The reviews on a product may be positive or negative, the negative reviews will attract the customers more than a positive review. These fake reviews can affect any business which leads financial profit or loses. Generally reviews will appear in e-commerce websites like jabong, Flipkart, Amazon etc. Due to the financial reasons many fake reviews will be appeared in these websites. The company owners will intentionally motivate some of the people to write the fake reviews to improve their business towards another product. Day by day every one focus the online reviews before planning for hotel, to buy a product, or to stay in any home stay.

Hence, people are not subject to physically inspect the product when buying online so they drastically unwantedly/wontedly depend on reviews of other buyers this must be made truthful as much as possible so that the buyer is not cheated with fake reviewers or spammers time and again. The problem is simple yet tiring to be accomplished through/read every review to mark it as a fake or ambiguous category this must be done systematically to get to the root of the problem. This problem can be addressed by training an ML model which deals with the review section to flag a particular review as genuine or spam. The interesting thing is spammers who didn't use the product can be caught this way. A spam review or the usage of different customer id can be used to filter review of the product falsely to get a good rating of the product. This can be filtered by checking the use of words like "awesome", "so good", "fantastic" etc. can be flagged. Since they tend to hype the product or they try to emulate genuine reviews with the same words using it again and again to make an impact on the buyer. Hence the issue of spam filtering requires huge data to train and be effective with added domain knowledge such as sarcasm sentences used by users to show their dissent towards the product, sometimes the product is good but not the delivery or the packing which affects the review classification. Here, an NLP technique is used to identify such reviews instead of misclassification to a negative review as in sentiment analysis. To remove unwanted or outdated product reviews those include data pre-processing.

Spam and person who write it are called review spammers. A review spam may be also referred as opinion spams, fake reviews, deceptive reviews, fraudulent reviews, or non-genuine reviews etc.

The focus of this research is to create an environment of online E-commerce industry where consumers build trust in a platform where the products they purchase are genuine and feedbacks posted on these websites/applications are true, are checked regularly by the company where the number of users is increasing day by day, henceforth companies like Twitter, WhatsApp, Facebook use sentiment analysis to check fake news, harmful/derogatory posts and banning such users/organizations form using their platforms. Parallel to that E-commerce (Flipkart, Amazon) industries, hotels booking (Trivago), logistics, tourism (Trip Advisor), job search (LinkedIn, Glass door), food (Swiggy, Zomato), etc. use algorithms to tackle fake reviews, spammers to deceive the consumers in buying below average products/ services. And the users need to be alerted of the spammer like "not verified profile" hence users need not worry about such false users.

So our Final aim in this is to implement best approach available for detection of fake reviews using opinion mining (sentiment analysis) techniques. To let users, know if each individual review is trustworthy or not for efficient use of money from users side.

CHAPTER 2

RELATED WORK

For fake review detection, there are numbers of machine learning algorithms. Using Machine learning techniques, fake reviews detection depends on behavioral features, linguistic and textual features and relational features.

1. A Framework for Review Detection-Issues and Challenges: Growth of Internet technology leads to the growth of ecommerce as well as associated review sites. Gradually people used to prefer online mode of marketing. Availability of millions of products and services on ecommerce sites makes it difficult to search for the best suitable product according to requirement. The best way to overcome this is to follow opinion of others who have already tried this product. A part of many independent review sites of few companies have their own review system. The huge amount of users generated content provided by these sites provides a lot of information regarding various products and services. While online reviews are useful but blindly trusting them leads to various dangerous consequences for both the seller as well as the buyer.

A lot of work is done on spam detection such as Email spam, SMS spam, Web spam, Social Media Spam, Search Engine Spam, and Video Spam. However Opinion spams are different and are found in product review sites. Opinion spams are highly implicit and pretend to be honest options voiced by actual users or customers. It is logically impossible to recognize fake reviews by simply reading it.

A Framework is developed to find a set of attributes to identify fake reviews some of them are Availability of labeled datasets, appropriate feature selection, early detection, cross site verification, how to deal with enormous data generated by review and ecommerce sites.

Disadvantages: This Algorithm needs labeled data and is only 90.19% Accurate for Supervised learning and 83.70% Accurate for Unsupervised learning

Solutions: More Research needs to be done, Big data Techniques can be explored, Early Detection

2. Detection of fake opinions on online products using Decision Tree and Information Gain: Now a day's people have started to give fake reviews on the products. The reviews on a product may be positive or negative, the negative reviews will attract the customers more than a positive review. These fake reviews can affect any business which leads financial profit or loses. Generally reviews will appear in e-commerce websites like jabong ,Flipkart, Amazon etc.. Due to the financial reasons many fake reviews will be appeared in these websites. The company owners will intentionally motivate some of the people to write the fake reviews to improve their business towards another product. Day by day every one focus the online reviews before planning for hotel, to buy a product, or to stay in any home stay. Now a days internet has become everything and it is too fast. People are interacting with the social network across the world because they can share all the information and their thoughts. Through the internet people can do the online shopping, human tendency to do survey on products which they want to purchase, lots of reviews on a particular products will be available on their company's website like flipkart , amazon . The online reviews can change decision of the customer and they can finalize a product by comparing with the different brands of products, the customer can select the product and satisfy their requirement only if the reviews are not fake. On the other hand if the reviews are fake then it misleads the customer.

Disadvantages: a small change in data can cause large change in structure of the decision tree causing instability, Information Gain is biased and does not hold good for larger value always.

Solutions: More Research needs to be done, more site comparison is required

3. Detection of fake reviews on social media using machine Learning algorithms:

The main idea of our research problem is to recognize the hidden patterns of fake reviews by using a clustering model based on cosine similarity among the reviews. We would like to emphasize that the objective is not using unsupervised learning to address the text classification problems but incorporating the result from clustering into the set of predictor attributes as an input to build the fake review classifiers. Our research underlined the importance of integrating the clustering step into data preprocessing. Although it was not significant, clustering can improve text classifying performance. By conducting separate reviews for each cluster, machine learning models can perform

better. Non-text features are truly significant in solving fake review problems. In this study, what we are concerned about is the trustworthiness of the reviews, thus we need a metric that can evaluate the credibility of reviewers. Yelp has done an excellent job in evaluating the reviewers by allowing their customers to assess the reviews and reviewers. The length of a reviewer's membership at the time of the reviews also demonstrated a significant impact on classifying fake reviews. We explained that fake reviewers usually create new accounts for their activities. Hence, we believe the future research needs to include the features to be obtained by tracking reviewers' activities and utilizes those features to measure the reviewers' credibility.

Disadvantages: it relies on one technique and user creating duplicate account cannot be avoided.

Solutions: More Research needs to be done, more site comparison is required

CHAPTER 3

OPINION MINING BASED FAKE PRODUCT REVIEW MONITORING AND REMOVAL SYSTEM

3.1 OPINION MINING BASED FAKE PRODUCT REVIEW MONITORING AND REMOVAL SYSTEM

We are going to explain how an opinion mining system works and also explain how it can be used to monitor and remove fake product reviews.

1) Dataset: Dataset used is "amazon academic review" which contains reviews, useful votes, ratings, user id, and many other attributes. The useful parameters are retrieved for feature engineering. The dataset contains thousands of original and fake reviews mixed to easily assess the accuracy of the model being implemented using this dataset. The Yelp dataset released for the academic challenge contains information for 11,537 businesses. This dataset has 8,282 check-in sets, 43,873 users, 229,907 reviews for these businesses (www.yelp.com/dataset). The dataset is challenging since it contains a large set of varied reviews and parameters for training any algorithm.

2) Pre-processing: Pre-processing is the first step in analyzing any dataset which includes removing unnecessary attributes, punctuations, stop words, missing words, redundant words , etc. to clean the dataset for training purposes. This ensures proper training of the model.

3) Feature Engineering: This function involves all the methods to remove unwanted information from the dataset it is also called data cleaning. This step is very necessary to find the gaps and the relationship between the different attributes (columns) and use them to draw valid conclusions. The libraries from the NLTK package is a bag of words used to construct a corpus of words. Term frequency, tokenizer, Stopwords functions are imported from OrderedDict. Stop words are removed and unwanted words like is, then, to, why, etc. which are not required in this context and do not add value to feature engineering are grouped under Stopwords coming under the English language. Term frequency counts the number of times a particular word has occurred and that can be used by spammers again and again to identify the spammer.

4) Sampling of data: Since a huge number of reviews are used in the dataset the data is

subjected to sampling before even fed to the classifier. The sampling is done to lower to weight on the classifier that loads the data in chunks. Here, different labels are used to authentic the fake reviews and then concatenate two columns after labelling and return the data frame.
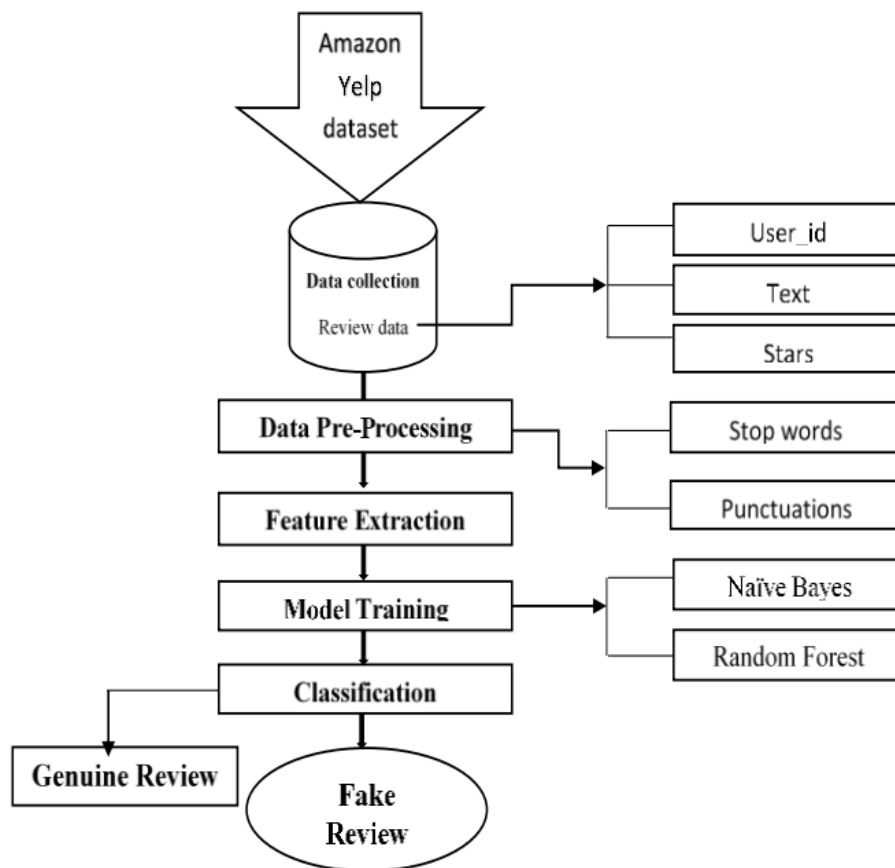

Figure 3.1: Model Diagram for fake review detection

Naïve Bayes algorithm:

A Naive Bayes calculation was utilized to assemble a double arrangement model that would anticipate if the survey's conclusion was positive or negative. A Naive Bayes classifier expects that the estimation of a specific component is free of the estimation of some other element, given the class variable. It utilizes the preparation information to compute the likelihood of every result dependent on the highlights. One significant trait of the Naive Bayes calculation is that it makes suspicions about the information. It expects that all the highlights in the dataset are

autonomous and similarly significant. The equations (1), (2), and (3) shown below are the standard form of any Naïve Bayes constituted problem, these are used to compute the probabilities for predicting values that are in the range (0, 1). Where p is a probability, a, b, xi , y, yi are values of which probability is calculated, σ is the standard deviation and μ is the mean of the attributes.

$$p\left(\frac{a}{b}\right) = \frac{p\left(\frac{b}{a}\right)p(a)}{p(b)}$$

$$Posterior = \frac{prior * livelihood}{excellence}$$

$$p\left(\frac{x_i}{y}\right) = \left(\frac{1}{\sqrt{2\pi\sigma^2_y}}\right)\exp\left(-\frac{(x_i - \mu_y)^2}{2\sigma^2_y}\right)$$

Figure 3.3: Naïve Bias Equation

Random forest classifier:

It is a supervised learning algorithm used to train and test machine learning models. The "forest" means an ensemble of decision trees trained with the "bagging" method. Here decision trees are combined to increase the performance and learning of the model to get good overall results. It basically merges multiple decision trees to amplify the performance of the random forest and get a more accurate prediction [13].

a) Accuracy= TP+TN/ FP+FN+TN

b) Precision= TP/TP+FP

c) Recall (sensitivity) = TP/TP+FN d) F1_score= 2*(Recall*Precision)/ (Recall + Precision)

All the above parameters determine the performance of the model; the results of models are shown along with the confusion matrix.

S. No	Parameter	Naïve Bayes (in %)	Random Forests (in %)
1.	Accuracy Score	79.007	**89.487**
2.	Precision Score	70.224	**85.577**
3.	Recall Score(Sensitivity)	99.099	94.389
4.	F1 Score	82.169	**89.768**

Figure 3.4: Model Diagram for fake review detection

From Table 1, It can infer that the two models performed fairly well except that the random forests classifier is better when compared. Hence random forests have got better accuracy, precision score, and F1 score. It is concluded, a random forest classifier can be used for the fake product review monitoring and removal approach. When compared to the models for diverse applications, they perform well in certain fields and incompatible in some areas, hence their application needs some experience.

3.2 DETECTION OF FAKE OPINIONS ON ONLINE PRODUCTS USING DECISION TREE AND INFORMATION GAIN

Now a day's online shopping is in leading, because of fake reviews the ratings of the branded products becoming down. The major task is to focus on identifying the fake reviews. The Decision Trees classifiers Technique is used in this work, reviews has been extracted and collected to identify the fake review by using six different conditions that is star ratings, Response, Reply, Useful Profile, Profile status, Template conditions. In this paper online reviews of the product are extracted by using Web harvy crawler. Potential features are extracted from set of collected reviews, then define the rules for Decision Rule Classifier to segregate the fake reviews by using decision Classifier technique based on several criteria. Final step is to identify the best feature to determine review is fake or not.
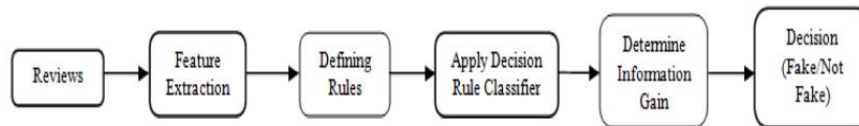
Figure 3.5: Block Diagram for proposed method

Following rules are defined to identify whether reviews are fake or not.

R1 (Response): It is to check whether the reviewer got any response for his review. Generally when a reviewer write a fake review on any particular product then company people will respond to the review quickly.

R2 (Useful profile): This is to check whether the reviewer profile is useful or not. If the reviewer profile is real profile then people will like the reviewer profile by stating "Use full information" .If the reviewer is giving the misguiding statements then will not like the profile.

R3 (Template): This is to check whether the reviewer uses standard Template or Not. If the review contains Template then definitely the review will be fake. If the reviewer wrote any negative or positive review within one sentence without stating the reason then it is fake.

R4 (Stars < 2): This is to check how much ratings the reviewer has given. Without any reason if the reviewer has given less than 2 ratings then it will be a fake review.

R5 (Reply): This rules is to check whether the reviewer has replied for the response from the company. If the reviewer is not a fake then definitely he will be replying for the response.

R6 (Thick): This rules is to check the thickness of the reviewer profile. The reviewer profile has full information (Thick) about his details which indicate reviewer is not a fake person else the reviewer profiles does not contain any detailed information (Thin), considered as fake review.
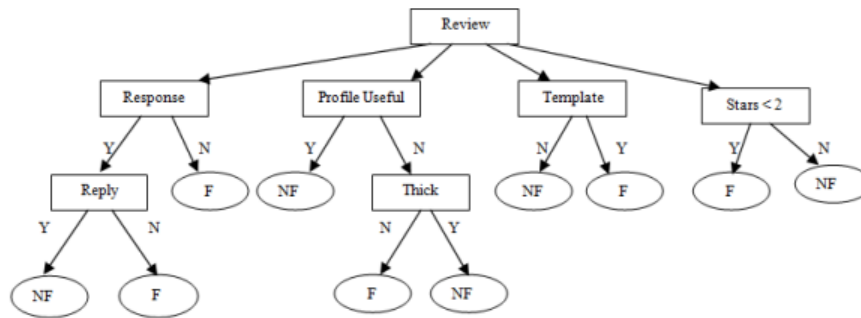
Figure 3.6: Machine Learning based Fake Review

Here: F=Fake, NF=Not Fake, Y=Yes, N=No.

Algorithm for decision rule classifier:

- Start

- Read the review and its features such as Response, Useful profile, Template, Stars ratings, Reply, Thick.

- If the response to the review is yes and there is no reply for response then it is a fake review otherwise it is not a fake review.

- If the reviewer profile is not useful and not thick then it is fake review.

- If the review uses template then it is a fake review otherwise it is not a fake review.

- If the star rating is less than 2 then it's a fake review otherwise it is not a fake review.

- Stop.

Huge data from the web will be collected using the WEB Crawler to get the online reviews. The web Crawlers extracts all the reviews from the web and stores in the form of text documents. In this work Web harvy tool is used for collection of data and experimentation. Dataset is prepared by collecting the reviews of the web users on the product called "LED TV" from the website "www.amazon.com",200 online reviews are collected for experimentation. From 200 reviews, 90 reviews has been detected as a fake reviews. Decision of the proposed approach is validated by manual review analysis, in which 94 reviews are detected as fake. Hence the success rate

of 96 % has been achieved as shown in the Table

Number of Reviews collected	200
Product	LED TV
Website	Amazon
Fake Reviews Identified Using proposed method	90
Fake Reviews Identified Manually	94
Success rate	96 %

Figure 3.7: Experimental results for fake reviews

Experimental result shows the efficiency of the proposed system. The Table shows the sample fake reviews of the product.

SL.No	Details	Review 1	Review 2
1	PRODUCT	LED TV	LED TV
2	REVIEWS	Not a worth product	Waste of Money.
3	STARS	1	1
4	RESPONSE	N	Y
5	REPLY	N	N
6	USEFUL PROFILE	1%	1%
7	PROFILE STATUS	Thin	Thin
8	TEMPLATE	Yes	Yes

Figure 3.8: Sample Experimental result of a fake review

The Proposed approach is evaluated using Entropy and Information Gain to identify the best feature to decide the fake review among six features.

Sl.No	Feature	Entropy	Information Gain
1	Response(R1) and Reply(R5)	0.3	0.3
2	Profile Useful (R2) and Thickness(R6)	0.3	0.3
3	Template(R3)	0.3	0.2
4	Stars < 2(R4)	0.3	0.2

Figure 3.9: Different accuracy measure for identification

Sample Illustration of Entropy and Information Gain: The above table shows the two methods to identify the best feature among these. The Entropy value will be same for all the measure here because of same leaf nodes at left and right. The Information Gain value is 0.3 which is High for Response, Reply and Profile Useful and Thickness and Information Gain 0.2 which is value is low for Template and Stars < 2 i,e 0.2 when compare to other. This Indicates that the Response, Reply and Profile Useful, Thickness will be the best feature to find out the accurate fake review. Information needed to specify the exact physical state of a system, given its macroscopic. Entropy is an expression of the disorder, or randomness of a system, or of the lack of information about it. The concept of entropy plays a central role in information theory. Entropy is degree of randomness of elements determined using the equation. Entropy, $E(x) = - \sum p(x) \log p(x)$ (1) Where $P(x)$ is the probability of x. Information Gain is nothing but Identification of best feature in the working set , that gives us less impurity. the amount of information gained about a random variable from observing another random variable. However, in the context of decision trees, the term is sometimes used synonymously with mutual information, which is the conditional expected value of the Kullback–Leibler divergence of the univariate probability distribution of one variable from the conditional distribution of this variable given the other one.

$$IG(x) \quad \sum_{ch=1}^{n}(E(x) - W * E(c$$

Figure 3.10: Entropy Gain and Information Gain Formula

Where,

E(x): entropy of node x W: Weighted avg.

E(child): Entropy of Childs. Information gain with highest value is

considered as best or most significant features in the identification of review is fake or not.

3.3 SURVEY ON FAKE REVIEW DETECTION USING MACHINE LEARNING TECHNIQUES

World Wide Web has drastically changed the way of sharing the opinions. Online reviews are comments, tweets, and posts, opinions on different online platforms like review sites, news sites, e-commerce sites or any other social networking sites. Sharing reviews is one of the ways to write a review about services or products. Reviews are considered as an individual's personal thought or experience about products or services. Customer analyzes available reviews and takes decision whether to purchase the product or not. Therefore online reviews are valuable source of information about customer opinions. Fake or spam review refers to any unsolicited and irrelevant information about the product or service. Spammer writes fake reviews about the competitors' product and promotes own products. The reviews written by spammers are known as fake reviews or spam reviews. Thus fake reviews detection has become critical issue for customers to make better decision on products trustworthy as well as the vendors to make their purchase.

The fake reviews are classified in two groups:

Untruthful reviews: These reviews promote or demote the products with positive or negative words respectively and misguide the customers. **Reviews on brands**: These reviews are not related to products, not on the different features of the product or services. Reviewer uses brand name repeatedly to promote a particular brand.

Jindal and Liu proposed three basic techniques for identifying fake reviews. These three approaches are as follows:

Review Centric Approach: This approach identifies review as fake review based on the content of reviews written by reviewers. In this method, various features like review content similarity, use of capitals, all capital words, use of numerals, brand name, similarity between products and reviews, repeated use of good and bad words in review. Reviewer

Centric Approach: This method depends on the behavior of reviewers. This approach considers information about users and all reviews that are written by them. Features used in this method are account age, profile picture, URL length, IP address, number of written reviews by one reviewer, maximum rating per day etc.

Product Centric Approach: This method mainly focuses on the product related information. In this method, sales rank of product, price of product etc are considered as features.

Initially fake review detection was introduced by Jinal et al. There are various ways to identify fake reviews. Machine learning technique is one of the ways to identify fake reviews. Machine learning model learns and make prediction. The basic steps involved in machine learning are data processing, feature extraction, feature selection, classification model generation. This process is shown in Fig.3.11
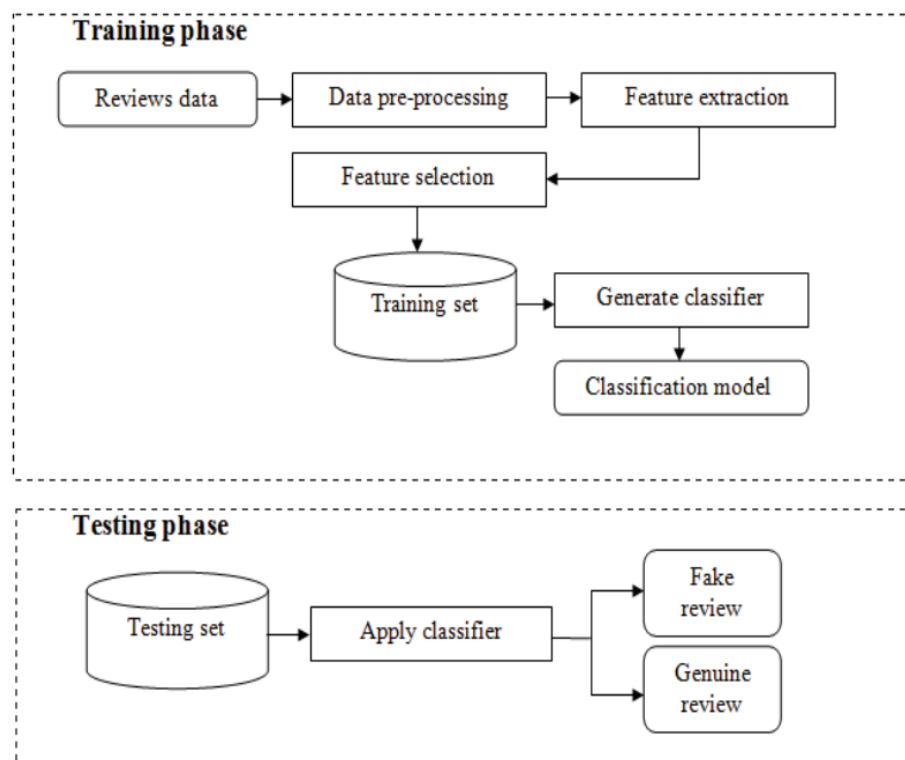
Figure 3.11: Machine Learning based Fake Review Detection

Machine learning approach for fake review detection works as follows:

Data collection: In this phase, review data will be gathered from various platforms like Amazon. These reviews could be for product or service

like hotel reviews.

Data pre-processing: In next step, data preprocessing is applied like punctuation marks removal, stemming, stop word removal etc. In punctuation marks removal, the whole text is divided into sentences, phrases or paragraphs. In the stemming process, stem will be created from every word in dataset. In stop word removal phase, frequently used group of words like determiners, articles and preposition will be detected and removed. After removing these words, only important words will be retained for the next step.

Feature extraction and selection: In this step, features are extracted from the preprocessed data. The different types of features which are used to detect fake reviews are classified as linguistic features, relational features, and behavioral features. The classification is shown in Fig. **Classifier model construction and testing:** For training purpose, small set of labeled data is used. In this phase, classification model is generated by using the training review dataset. The reviews used for this purpose are already labeled as fake or genuine review. Once the classifier is trained, it will be tested using test dataset. The different machine learning algorithms which can be used for model construction are naive bayes classification, decision tree algorithm, support vector machine, k-nearest neighbor, logistic regression, etc. The performance of fake review detection method depends on labeled data used for training purpose, correct selection of features and data mining techniques used for detection.

CONCLUSION

Identifying fake reviews from a large dataset is challenging enough to become an important research problem. Business organizations, specialists and academics are battling to find the best system for opinion spam analysis. A single algorithm cannot solve all the problems' and challenges faced in today's generation with advancements in technologies, though a few are very efficient in analysis. More future work and knowledge is needed on further improving the performance of the opinion spam analysis, and developing one that is consistently efficient across all categories of data. In future work, hybrid models and new models can be tried for the fake review detection model. By using Google co-lab and NVIDIA graphics GPU, the research can speed up the process of execution. The methods described here have the most efficient and effective methods used to detect fake reviews. It should be noted that the least efficient methods are the most commonly used. This is because we are interested in the most effective methods and not the most efficient. The most important part of an algorithm is its efficiency. Efficiency is not just about execution time. The efficiency of an algorithm is about the time taken for training the model and the time taken for the prediction.

REFERENCES

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

[2] Jitendra kumar Rout, Amiya Kumar Dash and Niranjan Kumar Ray "Framework for Fake Review Detection: Issues and Challenges" IEEE Xplore ISBN:978-1-7281-0259-7/18 2018.

[3] Nidhi A. Patel and Prof. Rakesh Patel "A Survey on Fake Review Detection using Machine Learning Techniques" IEEE Xplore ISBN: 978-1-5386-6947-1/18 2018.

[4] Sanjay K.S and Dr.Ajit Danti "Detection of fake opinions on online products using Decision Tree and Information Gain" IEEE Xplore ISBN: 978-1-5386-7808-4 2019.

[5] Syed Mohammed Anas and Santoshi Kumari "Opinion Mining based Fake Product review Monitoring and Removal System" IEEE Xplore ISBN: 978-1-7281-8501-9 2021.

APPENDIX

Slide 1:

OPINION MINING BASED FAKE PRODUCT REVIEW MONITORING AND REMOVAL SYSTEM

Shivathmaj Shenoy M
4CB18CS085

Slide 2:

Agenda

- Overview
- Working of Fake Review detection
- Fake Review detection Using Naïve Bias
- Fake Review detection Using Random Forest Classifier
- Fake Review detection using Decision tree
- Fake Review Survey Using Machine learning Approach
- Scope of Growth and Disadvantages
- Conclusion
- Reference

Slide 3:

Overview

- The reviews on a product may be positive or negative, the negative reviews will attract the customers more than a positive review. These fake reviews can affect any business which leads financial profit or loses.
- Due to the financial reasons many fake reviews will be appeared in these websites. The company owners will intentionally motivate some of the people to write the fake reviews to improve their business towards another product.
- The focus of this research is to create an environment of online E-commerce industry where consumers build trust in a platform where the products they purchase are genuine and feedbacks posted on these websites/applications are true
- So our Final aim in this is to implement best approach available for detection of fake reviews using opinion mining (sentiment analysis) techniques. To let users, know if each individual review is trustworthy or not for efficient use of money from users side.

Slide 4:

Working of Fake Review detection

1. Preprocessing
2. Dataset
3. Feature Engineering
4. Sampling of Data

Slide 5:

Working of Fake Review detection

Figure 3.1: Model Diagram for fake review detection

Slide 6:

Fake Review detection Using Naïve Bias

$$p\left(\frac{a}{b}\right) = \frac{p\left(\frac{b}{a}\right)p(a)}{p(b)}$$

$$Posterior = \frac{prior \cdot likelihood}{excellence}$$

$$p\left(\frac{x_i}{y}\right) = \left(\frac{1}{\sqrt{2\pi\sigma^2}_y}\right)\exp\left(-\frac{(x_i-\mu_y)^2}{2\sigma^2_y}\right)$$

Figure 3.3: Naive Bias Equation

Naïve Bias Vs. Random Forest Classifier

S. No	Parameter	Naïve Bayes (in %)	Random Forests (in %)
1.	Accuracy Score	79.007	89.487
2.	Precision Score	70.224	85.577
3.	Recall Score(Sensitivity)	99.099	94.389
4.	F1 Score	82.169	89.768

Figure 3.4: Model Diagram for fake review detection

Fake Review detection using Decision tree

Figure 3.6: Machine Learning based Fake Review

Fake Review Survey Using Machine learning Approach

Figure 3.11: Machine Learning based Fake Review Detection

Fake Review Survey Using Machine learning Approach

- World Wide Web has drastically changed the way of sharing the opinions. Online reviews are comments, tweets, and posts, opinions on different online platforms like review sites, news sites, e-commerce sites or any other social networking sites
- 1. Untruthful reviews
- 2. Reviews on brands
- 3. Review Centric Approach
- 4. Centric Approach
- 5. Product Centric Approach

Fake Review Survey Using Machine learning Approach

- Data Collection
- Data pre-processing
- Feature Extraction and selection
- Classifier model construction and testing

Application:
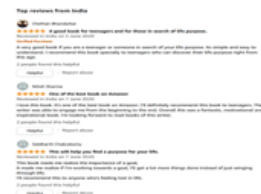
Scope of Growth and Disadvantages

- No Fraud Algorithm we need to choose a Algorithm based on Criteria and this is a manual task
- In future work, hybrid models and new models can be tried for the fake review detection model.
- Process is slow we can increase its speed by using Graphical Processing Unit
- Need for constant evolution of methods because humans always find smarter ways to cheat the system

Conclusion

- Identifying fake reviews from a large dataset is challenging enough to become an important research problem. Business organizations, specialists and academics are battling to find the best system for opinion spam analysis.
- The most important part of an algorithm is its efficiency. Efficiency is not just about execution time. The efficiency of an algorithm is about the time taken for training the model and the time taken for the prediction

Reference

- [1] Dhairya Patel, Aishwarya Kapoor ans nd Suneet Kommissar "Fake review is detected on using opinion mining" International Research journal of Engineer ting and technology (IRJET) , volume 8, Issue 07,Dec 2016.
- [2] Sterdho kumar Rout, Aditya Kumar Dash and Niranjan Kumar Ray "Framework of the Fake Review Detection: Issues and Challenges" IEEE Xplore 9BN4578.1.7281-8795-73 6   2016.
- [3] Nidhi A. Patel and Prof Rakesh Patel "A Survey on Fake Review Detection using Machine Learning Techniques" IEEE Xplore ISBN : 978. 3. 4306. 6847. 0061. 2016.
- [4] Sanjay K.R and Dr Ajit Danti "Detection of fake opinions in a real time predicament using Decision Tree and Information Gain" IEEE Xplore ISBN : 978. 3. 4306. 7686.4 2019.
- [5] Syed Mohammed Anas and the model Kumar "Opinion Mining based Fake Product review Monitoring and Removal System" IEEE Xplore ISBN : 07.8.1. 7281.8481.8 2021.

THANK YOU