Fake Review Detection Using Machine Learning Algorithms

Ajay Rana, Harsh Shah, Vinay Gor
School of Systems and Enterprises
Stevens Institute of Technology
Hoboken, New Jersey, USA

Abstract: New sellers on Amazon are trying to make their products appear on the top searches using amazon choice feature which can influence buyer and thus increase in sales for that seller. The seller posts multiple ads on different social media platforms like Instagram and Facebook for users to be part of beta tester. The interested public will click on the ads, and a deal takes place between the seller and the buyer where the buyer must post a 5-star review in return for a free product from Amazon. The importance of reviews is for companies to make design decisions of their product and services. However, these days reviews are misused; for example, 5-star reviews push the product to feature on the first few pages of the Amazon product list and entice other buyers to purchase based on the review. We propose a fake review detection system that detects fake reviews using a combination of vectorization and machine learning models like Support Vector Machine and Multinomial Naïve Bayes. For validation, we tested fake reviews posted by us on Amazon, resulting in 71% accuracy.

Keywords—Fake review detection, Amazon fake review detection

I. Introduction

Since Amazon's early days, customers have relied on reviews to determine the quality and authenticity of the online marketplace's products as they are the one significant metric. Amazon's listings often have thousands of reviews instead of handful found on competing marketplaces. However, many of those reviews cannot be trusted. Many of the reviews cannot be trusted. UCLA and USC released a study that found more than 20 fake reviews related to Facebook groups with an average of 16,000 members. In more than 560 postings each day, sellers offered a refund or payment for a positive review, usually around $6[H1]. Thousands of fake reviews have flooded Amazon, Walmart, eBay, and others, as sales have skyrocketed. In this project, we will explore these fake reviews' essence, create multiple models to predict whether a review is genuine or not, and finally validate the model by checking any Amazon product reviews against the model.

There is no universally accepted definition of a fake review. Hu et al. define review fraud as the act of vendors, publishers, writers, or any third-party monitoring online reviews and posting nonauthentic online reviews as real customers to boost product sales. Fake reviews in the past were done by the company employees there family friends and members. Most of these are purchases were unverified purchase as Amazon do not verify if the product was bought in actual and then review was done. Considering that fake reviewers have not used the product, write general reviews just positive and just praising the product how good it is in terms of looks and feels.

In recent times it has become hard to catch fake reviews as they are verified and more detailed. People are from the general public who click on ads that show free products, and once an ad is clicked, deals are made with the seller in chat. People buy products from Amazon, and once they write a positive and five stars rated review. They are refunded the money through a digital wallet. These reviews may seem genuine but, in general, fake. Even if people liked the product after purchase, this is not a fair practice.

II. Methodology

There is no proper method or algorithm to verify if the review is fake or not, but few methods for detecting a fake review have been developed with good overall accuracy.

What comprises a fake review in general:

- **Five-star Rating:** Fake reviews have a high rating as the seller's motive is to attract customers by high Rating as this the first thing a customer looks for while searching for a product.
- **Positive Review:** Fake reviews are generalized and have mostly positive things to write without discussing a product's working or specifications.
- **Type of Purchase:** Fake review purchases are mostly unverified as they would likely not buy a product to review.

- **Length of review:** A fake review tends to be short as no person would invest his time to write such a lengthy review.
- **Words usage:** fake reviews tend to use similar words such as love, great and good in more frequency than genuine reviews.

A. Data Exploration

In this section, we try to explore our dataset. We have an Amazon reviews dataset that has reviews already classified as fake or real. Dataset contains equal real and fake reviews for better modelling of the dataset. We will compare how real and fake reviews differ in Rating, type of purchase, and overall review sentiment.

Label '0' means fake reviews, and label '1' means real reviews. Since the data is equally divided, the number of real and fake reviews of each Rating are almost equal in number.

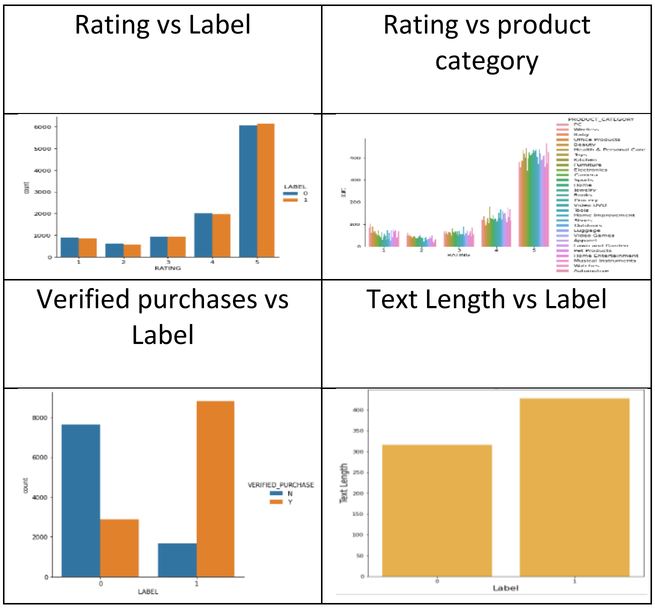

Fig. 1 Comparing different variables with Rating and Label

It can be seen from the above figure that most of the fake reviews have Unverified purchases by a large amount, and similarly text length of these reviews is short compared to genuine reviews. However, Rating is the same for both the type of reviews and trend continues in all the product category.

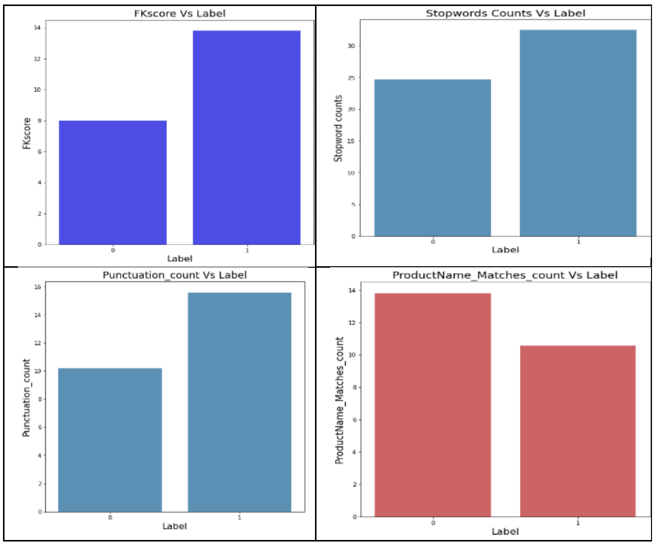

Fig.2 Analysis of Review Text to Label

In fig.2 text of the review has been analyzed as to how it varies from each other. The number of punctuation in the fake review is less than the genuine review, which shows the quality of the review being written. The review's readability can be measured by the Fk-score, which tells us how readable the review. Fake reviews have an Fk -score of 8, which means that 8th-grade students can read it, whereas genuine reviews have an Fk-score of 13, a high readability score.

Review text was also checked for the number of stopwords, and it was also checked whether the review text matched the product title.

The word cloud below of both reviews shows us that fake reviews are similar to those in the fake reviews. However, the frequency of words used is far different, which tells that the real reviews are far more passionate and descriptive than the generic words used to express quality in the word cloud of fake reviews. In the fake review, word cloud, it can be seen that it reviews talks about only how good the product is. For example, the size of words like great, really, and love is far more prominent in fake reviews than genuine reviews.

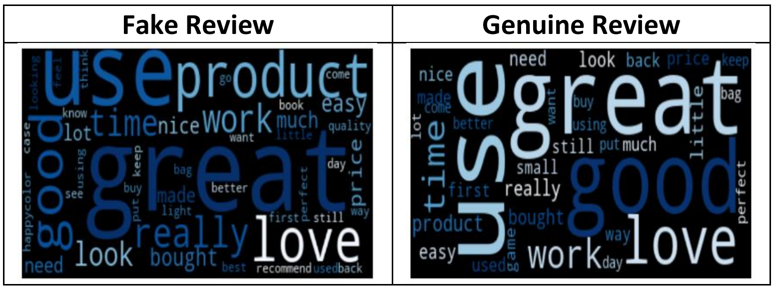

Fig.3 fake review and real review word cloud

Sentiment analysis:

The sentiment of each review was calculated, and the average score was calculated. It can be seen that fake reviews have a more positive sentiment than its counterpart, which shows that they tend to praise the product more in general than have a Neutral approach, which can be seen in real reviews.

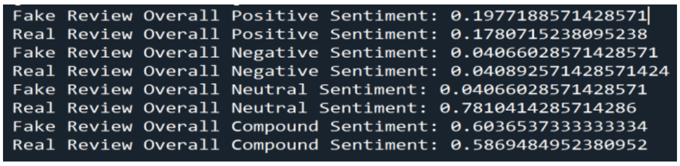

Fig.4 Average Sentiment score of reviews

B. *Modelling*

TF-IDF

Term Frequency - Inverse Document Frequency is a technique to evaluate the importance of words in a document is a collection of the corpus. This is done by assigning weight to the words in a document, the weight of the words is directly proportional to the number of times words appear in the corpus. This technique also offsets the common words appearing in the document giving them the low weightage and rank as they do not mean in the corpus [4].

TF

Frequency evaluates the number of times a word/words appeared in a document. As the Length of the document can be different hence frequency cannot be the only measure of importance. So the term frequency is considered which is usually frequency divided by the number of words in that document.

$$TF = \frac{Number\ of\ times\ \ t\ appear\ in\ a\ document}{Number\ of\ words\ in\ a\ document}$$

IDF

IDF evaluates the importance of words in a document. While TF gives equal importance to all words TF-IDF looks for words which are important to the document and gives less importance to words such as "of", "is" etc. thus weighing down frequent words and scale up the importance to rare words.

$$IDF = \log\left(\frac{Number\ of\ documents}{Number\ of\ documents\ with\ words\ in\ it}\right)$$

Finally,

TF-IDF = TF x IDF

All words are assigned as vectors and given values between 0 & 1. Where more weighted words are high valued and less weighted words are valued low.

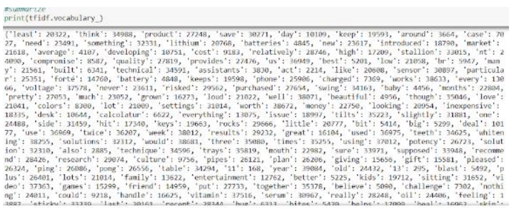

Fig.5 TF-IDF Final weighted words

SVM

Support Vector Machine or Support Vector networks are supervised machine learning models used for classification & regression. SVM classifies the vector categories in a hyperplane with a maximum distance between both categories separated by equal distance [5].

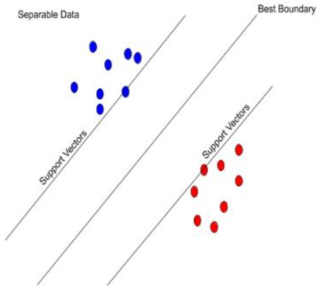

Fig.6 Support Vector Machine modelling

Support vectors are given by,

$$f(x) = \sum_i \alpha_i y_i (x_i^T x) + b$$

Where,

separable data x_i labelled in two categories.

$y_i = \{1, -1\}$ to find a weight vector w for $i = 1,2, \ldots, N$

Multinomial Naïve Bayes

The multinomial Naive Bayes classifier is suitable for classification with discrete features (e.g., word counts for text classification). The multinomial distribution usually requires integer feature counts. However, in practice, fractional counts such as TF-IDF may also work [H3].

Multinomial Naive Bayes is a specialized adaptation of Naive Bayes that is planned more for content reports. Though straightforward Naive Bayes would demonstrate an archive as the nearness and nonappearance of specific words, multinomial Naive bayes unequivocally models the word checks and alters the fundamental calculations to bargain within [H2].

To understand how Naïve Bayes works, first, we must understand the concept of Bayes' rule. This probability model was formulated by Thomas Bayes and can be written as:

$$Posterior\ Probability = \frac{Conditional\ Probability\ *\ Prior\ Probability}{Predictor\ Prior\ Probability}$$

$$P\left(\frac{A}{B}\right) = \left(\frac{P(A \cap B)}{P(B)}\right) = \frac{P(A) * P\left(\frac{B}{A}\right)}{P(B)}$$

where,

$$P(A) = \ the\ prior\ probability\ of\ occurring\ A$$

$P\left(\frac{B}{A}\right)$ = the conditional probability of B given that A occurs

$P\left(\frac{A}{B}\right)$ = the conditional probability of A given that B occurs

$$P(B) = the\ probability\ of\ occurring\ B$$

The (conventional) Multinomial Naïve Bayes model considers a document D as a vocabulary-sized feature vector x, where each component x_i is the check of term i in document D. This vector x at that point takes after a multinomial dispersion, driving to the characteristic classification work of Multinomial Naïve Bayes.

$$p(x \mid C_k) = \frac{(\sum_i x_i)}{\prod_i x_i!} \prod_i P_{ki}^{x_i}$$

But in this case, we use TF-IDF and when utilizing TF-IDF weights rather than term counts, our feature vectors are (most likely) not taking after a multinomial dispersion any longer, so the classification work is not hypothetically well-founded any longer. Be that as it may, it does turn out that TF-IDF weights rather than count work much superior.

III. RESULTS

The following models were implemented for comparison and alongside are the results:

Model		Precision	Recall	F1 score
Support Vector Machine		70.23	70.14	70.11
Multinomial Naive Bayes		69.65	69.42	69.33
Random Forest		59.98	59.69	59.04
Decision Tree		56.83	56.82	56.82

For validation we tested a few fake and genuine reviews. We considered a fake and a genuine review for sponsored products found on social media platforms, where the seller offered refund in exchange of a 5-star review. The genuine review was a 3-star review extracted from amazon.com, while the fake review was a self-written 5-star review.

Example 1:
Product: Projector

Fake review:
"The picture quality and clearity of this projector is beyond words. I am glad I took this projector instead of tv. It works really well. The remote control with it also works good. One of the interesting thing is they also offer hdmi cable with the projector."
Genuine review:
"this $200 dollars projector is already cheaper than any of the other generic kind of projectors (but not cheap enough because is $200 still). Or i can say the cheapest in the market. Don't get me wrong i like this projector, but am a power user and an A/V specialist so i have quality equipment and i know the difference and have experience and seen them most if not all: PROS: -Cheapest in the Market for a Home Cinema Projector. -relative 1080p but not actual 1080p but better than 720p -very nice speaker in treble and mid range -very good presentation projector -has 2 HDMIs, 2 USBs, SD Card slot, AV input port, Audio Out Port or Headphones -3 years limited warranty CONS -Seems disposable (assuming the 50,000 hrs lamp life is real. Scenario: let us say you use the Projector for 12 hours in a day so for 365 days /year equals to 4,380 hours) Approximately the Life expectancy of this projector is 11 years. BUT that is TBD. -Huge and bulky projector. -NOT 1080p native resolution -NOT 6500 Lumens -very limited features or programming -unlike other high end projectors you cannot replace the lamp/bulb yourself in this projector MY VERDICT: Still very good Projector in its own CLASS, very good $200 alternative than those high end Projectors ranging $800-$5k. Hopefully they can make cheaper technology with a very competitive features and specs."
Results:
Both the SVM and MultinomialNB classified the fake and genuine review correctly.

Example 2:
Product: Watch

Fake review:
"I was looking for a black affordable smart watch. This watch really seeked my attention. It is worth buying and has good looks. The dial is really big and looks perfect professionally and casually as well. Even though dial doesn't have many options the ones which are available are decent enough. It perfectly tracks all activities."
Genuine review:
"The watch is good, I use it to check BP/HR/ for sleep overall my only complaint is that the tracking for steps is not correct there is no replacement for the band which sucks"

Results:
The fake review was classified correctly by both the models but the genuine review was classified as fake by MultinomialNB model

Example 3:
Product: Pillow
Fake review:
"I always had a pillow which is firm. I ordered this pillow because of the firmness, shape and size. Also this is orthopedic pillow. I really liked the pillow it's shape it's size and it's firmness. Enjoying this pillow. Not only this can be used for sleeping but I also use it for back while studying and has been pretty much relaxing. It's softness and shape has been consistent. Worth the purchase".

Genuine review:
"Purchased this after reading multiple good reviews. Unfortunately for me this didn't work well at all. It's a couple inches too short so my neck kinks and muscles pull and spasm as a result. I figured it was an issue with my neck so I brought it to my chiropractor appointment to find out if it's me or the pillow. He suggested I measure from my neck to my shoulder and find a pillow that the height is the same. Even when I lie ony back, the pillow is way off and creates additional discomfort. Such as burning muscles around my head. Jaw problems, pain and tightness. And I always end up with a kink in my neck. I really wanted to love this pillow. Then I would've settled on simply being able to sleep a little better but it's made everything worse. To attempt to compensate for the lack of height I've lined my bed with a folded towel. Didn't help. Finally I found a really thin soft pillow, which helped a little. I hate how much money I loose to pillows that are either completely terrible or not for my needs, that simply don't make me any more comfortable but more uncomfortable. I think this would be great for a youth. Because I'm a 5' tall adult and it's far too small."

Results:
In this example both models classified the fake review correctly while the genuine review was classified as fake by both the models.

Similarly, we tested ten other pairs of fake and genuine reviews, out of which 8 pairs of the review were classified correctly by the SVM model while the MultinomialNB classified 7 pairs of review correctly.

IV. CONCLUSION

This paper performed an in-depth examination of supervised learning for fake review detection utilizing Support Vector Machine and Multinomial Naive Bayes utilizing TF-IDF, achieving classification accuracy of 70.14% and 69.42%, respectively, with bigram highlights and adjusted information. The prediction of the reviews are solely based on algorithms, and these are not 100% accurate at predicting if reviews are fake or not, but it can give us insight on how a fake review might be. As a reviewer might have liked product and do not have time for long reviews and write a short review and gives a 5-star rating might be labelled as a fake review.

REFERENCES

[1] Amazon is filled with fake reviews, and it's getting harder to spot them. https://www.cnbc.com/2020/09/06/amazon-reviews-thousands-are-fake-heres-how-to-spot-them.html

[2] Multinomial-Naïve-Bayes-Explained. https://www.mygreatlearning.com/blog/multinomial-naive-bayes-explained/

[3] MultinomialNB_scikit-learn. https://scikit-learn.org/stable/modules/generated/sklearn.naive_bayes.MultinomialNB.html

[4] TF-IDF http://www.tfidf.com/

[5] The SVM Classifier https://www.robots.ox.ac.uk/~az/lectures/ml/lect2.pdf