Fake Google Reviews in Alaska Using Data Mining

Project Final Report

Nick Sietsema
sietsemn@colorad.edu

Rebecca Toland
reto6656@colorado.edu

Amber Tin
amti9774@colorado.edu

Abstract

Google reviews has become one of the most important platforms consumers use to determine whether or not they will visit certain businesses. Consequently, this means an increase of positive or negative fake reviews to skew the rating of a particular business. It can be difficult for consumers to know if a review is honest based solely on the information available, such as ratings and customer feedback. In this paper, we present an in-depth analysis of fake Google reviews by looking at the reviewer centric approach through data mining applications.

Through a large-scale dataset of Alaska Google reviews, we dive deep into the characteristics of those fake reviews, and remove such reviews to get a better insight on what the actual rating of businesses would be. Can machine learning algorithms effectively help differentiate between fake and real Google reviews? To further explore this topic, we use the K-means clustering algorithm exploring the reviewer-centric model.

1 Introduction

Reviews are heavily used for purchasing decisions and those that are fake can be detrimental for customers and businesses. According to a 2016 study by the Pew Research Center, 82% of U.S. adults say they at least sometimes read online customer reviews before purchasing items for the first time, and 40% of that group say they almost always or always do so. Businesses can purchase fake reviews to improve their ratings, search results, and sales.

On Google, the accessibility to write a fake review is as easy as creating an account and posting it which goes public instantly upon submission. For businesses, they need to go through a Google review process to appeal the fake review due to Google being a third-party. From the article “How to Spot (&Remove) Fake Google Reviews”, 62% of consumers have said to have read fake reviews last year. The headache fake reviews have caused consumers and businesses has escalated the process to flag them in order to get the best possible rating for big or small businesses.

Although the FTC is cracking down on review fraud, artificial intelligence tools threaten to compound the problem by cranking out far more fake reviews that appear highly authentic. Therefore, efforts to eliminate fake reviews is imperative in order to protect customers and support honest businesses.

The purpose of our project is to detect and classify online reviews into real or fake through feature engineering and rule based classification.

2 Related Work

Several research studies have been conducted on the detection of fake Google reviews using data mining techniques.

In the study conducted by [Pendyala 2019], an extensive investigation was undertaken to address similar fake reviews, particularly through the Google platform. The research used many techniques including Naive Bayes, Linear SVC, Support Vector Machine (SVM), Random Forest, and Decision Trees algorithm. To enhance the

accuracy even further, features like the sentiment of the review, verified purchases, ratings, emoji count, product category with the overall score were used. With implementing these methods, this study was able to get a success rate of over 79%. The methodologies to this research were strong but for future enhancements to this existing body of knowledge would be to use real time based datasets.

An algorithm that could read real-time information could potentially flag incoming fake reviews, preventing it from even being posted to the platform and save the time of all users. Another item this research could have used to create a stronger case would be using logistic regression to be able to compare the attributes and find the best fit line between the two. Overall, the creative idea on using attributes, emoji count, verified purchases, rating sentiment of the review, and product category led to a strong success rate which will help the existing body of knowledge.

In the research conducted by [Hossain 2019], various algorithms were carried out to detect fake reviews including Support Vector Machines (SVM) with stochastic gradient descent (SGD) learning, Multinomial Naive Bayes, and Multi-Layer Perceptron with one hidden layer (MLP1) and two hidden layers (MLP2). The research used multiple approaches including supervised learning, semi-supervised learning and unsupervised learning. Having supervised learning is always controversial due to the fact of one being biased which could limit the results. Overall the research conducted by [Hossain 2019] led to some spectacular results following another researcher's approach [Ott] but with a slightly different approach. We think it's important for others to validate the research of others and expand upon it. This research has led to a stronger body of knowledge on the topic by giving an 89% accuracy on the semi-learning approach.

[Salminen, Kandpal, Kamel, Jung and Jansen 2022], found that using large language models (LLMs) to detect fake reviews had a near perfect success rate in comparison to human reviewers, which had a success rate of around 50%. An interesting approach of this research was using ChatGPT, RoBERTa (their own version of ChatGPT), and a baseline algorithm in NPL with a support vector machine(SVM).One big advancement to the field was the release of their dataset, model parameters, model weights, and initialization seed for experimentation reproducibility which allows users to build on their research. Other research has not released their results publicly so this research was generous to set an example as we are all working towards the same goal.

[Ott, Choi, Cardie, and Hancock 2011] studied a different perspective than the others mentioned in this section. They primarily focused on using genre identification, psycholinguistic deception detection, and text categorization. With these approaches, they introduce techniques including Naive Bayes and Support Vector Machine classifiers. The success rate of these combined features led to an accuracy of nearly 90%. Their research used automated classifiers by deeply researching different aspects of real and fake review. They created a large dataset of 800 opinions in which they developed automated deception classifiers from that information. The data cleaning this research used was the non-5-star reviews, non-English reviews, reviews with fewer than 150 characters, and reviews from first time users. The only problem with taking out the non-5-star reviews could potentially take away real reviews from the research causing a skewed result. Another issue with this is the small dataset used to determine whether a review is deceptive or not. In total, the results from this research was significant as they developed the first large-scale dataset containing a gold-standard deceptive opinion scam. They showed how deceptive opinion spam is beyond

the human eye, who judge at-chance. The findings have brought several theoretical contributions others are able to work off of and deepen the knowledge in this field

3 Dataset

The dataset our project is based on is Alaska Fake Google Reviews from Datarepo[1]. The dataset comprises an extensive collection of reviews, over 1 million. From the original untouched dataset, each review entry is characterized by many attributes including user id, name, business id, text (text of the review), pictures, time, and response from the business.

For data distribution, we looked at reviews at multiple avenues. First we looked at reviews with no content which consumed 411,557 reviews, as a result these were dropped from the dataset. After dropping these reviews with no content, we were left with 621,209 reviews. We were curious as to how many of these reviews were duplicates so we investigated further into just the unique reviews which consisted of 571,540 reviews. That means around 50,000 reviews were duplicates.

Next we looked at how many businesses were in the dataset and it comprises 12,689 different businesses. There was nothing else here that compromised the dataset quality.

The third thing we wanted to look at was the unique user IDs, which came out to be 278,478. This did not raise any red flags.

The next thing we looked at the time each review was posted since we thought it would be interesting to see if there were duplicated times the reviews were posted.

Looking at the timestamps, there was about a 300 difference between the regular dataset and a unique timestamp. This raised suspicion to fake reviews or was it a technological error.

The last thing we looked at was the rating distribution within the dataset. Below is the distribution of the rating across all the reviews.

3.1 Ratings

a. 5 star reviews: 651555
b. 4 star reviews: 201298
c. 3 star reviews: 85478
d. 2 star reviews: 62473
e. 1 star reviews: 21962

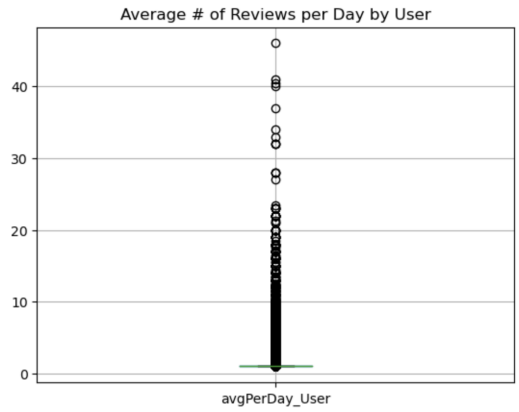
4 Main Techniques Applied

The data cleaning we utilized in our research consisted of removing noisy data, such as duplicate reviews which had well over 400K, the same user id, business id, rating, and timestamp Duplications such as same user id and business id, but different timestamps were retained for our analysis. We removed irrelevant attributes such as business response and pictures.

Our primary technique for classification involved feature engineering and unsupervised learning through k-means clustering. This step commenced after extensive cleaning and transformations to the original dataset, including dimensionality reduction. We analyzed the attributes to find correlations, such as ratings and the word count for reviews. However, we didn't find any strong correlations, especially with variables driving the ratings. The challenge is that fake reviewers can post both positive and negative reviews. Negative fake reviews can originate from companies who want to weaken their competition. Also, fake reviewers sometimes post four stars instead of five in order to avoid detection.

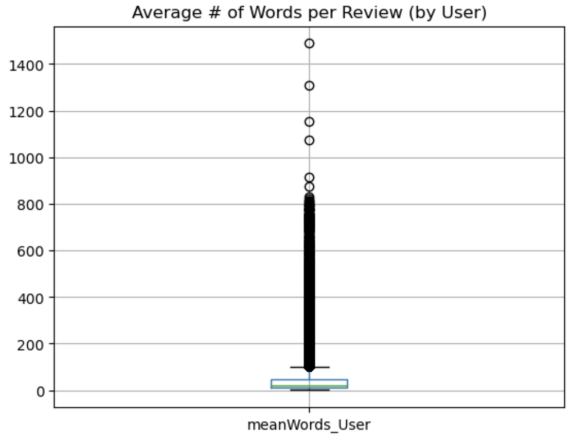
[1] https://datarepo.eng.ucsd.edu/mcauley_group/gdrive/googlelocal/

Ultimately, we found strong candidates to use for our model through analyzing the data at the reviewer level. We created a new attribute for the average number of reviews per day by user. This uncovered outliers that pointed to fake reviewers.



Average Number of Reviews by User

The number of words in a review can also provide insight into authenticity. Real reviewers had an actual experience with the product or service that they can describe in detail, hence using more words in their reviews. Therefore, we added an attribute for the mean words per review by user.



Average Word per Review by User

Below are the summarized results for these two features. This highlights the outliers and possible bot activity in online reviews, with a maximum average of 46 reviews per day by one user. This is a large deviation from the mean of one review per day. The mean number of words per review (38) was higher than expected and segregated those leaving few or even zero words in their comments.

	Reviews per Day	Words per Review
Mean	1	38
50%	1	22
Maximum	46	1,489

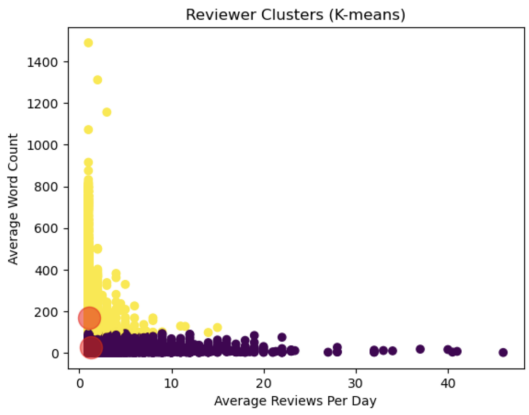
Feature Engineering Results

Next we used sklearn to develop the k-means clustering model, using the average reviewers per day and average number of words per review as the two dimensions. We also tested other features for the model, which didn't lead to meaningful results. Two clusters were formed to segregate real versus fake reviews.

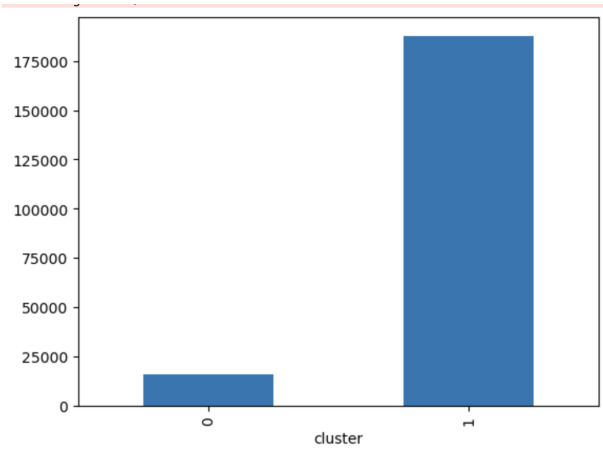
5 Key Results

Using unsupervised learning (specifically the K-Means algorithm), we were able to classify reviews into fake and real.

As can be seen in the cluster chart below, our model produced two distinct clusters that segregated potentially fake reviews (purple cluster) from real reviews (yellow cluster). Our model suggests that fake reviews tend to come from users who post many reviews per day with low word counts. The fake reviewers cluster comprises 8% of the data points, while 92% were classified as legitimate reviewers.



K-Means Reviewer Clusters



Cluster Bar Chart

The bar chart above shows the number of reviewers that were classified as fake (cluster 0) and those that are more likely to be legitimate (cluster 1)

Our data is consistent with previous research on the topic. For example, according to a CK Solutions report, 10.7% of Google reviewers are estimated to be fake. Our result of 8% is close to this estimate

6 Tools

For tools we used a number of different packages within the Python interface along with different communication means.

6.1.1 NumPy proved to be an invaluable tool, allowing us to utilize the np.where() feature to construct novel conditional arrays for key attributes such as fake names, mean word count, identifying users exhibiting multiple aliases, and flag user IDs associated with an average rating exceeding 3. The employment of NumPy significantly increased the efficacy of our K-means data mining methodology.

6.1.2 Pandas was a monumental help with our data cleaning and preprocessing of the large dataset. It allowed the diverse mathematical operations across heterogeneous arrays. It facilitated the decompression of the JSON file, giving access to its contents. Pandas also proved the partition of our data into workable subsets. Subsequently it allowed us to reintegrate the subsets back into a singular array.

6.1.3 The data visualization aspect of our project utilized the Matplotlib library. The tool

facilitated the construction of our important data mining techniques using K-means. Matplotlib's versatility empowered our project by bringing diverse dimensions of our dataset. It enabled us to create distinct histograms for each attribute, thereby revealing further exploration of certain topics.

5.1.4 Datetime proved to be helpful but changing the unix time into a readable time any user could easily interpret. This allowed us to run different commands and get an exact measurement of readable time.

5.1.5 Sklearn yielded significant advantages to our framework such as the cluster model and preprocessing for StandardScaler. The cluster implemented a seamless K-means clustering algorithm which was a cornerstone to our data mining technique.

7 Applications

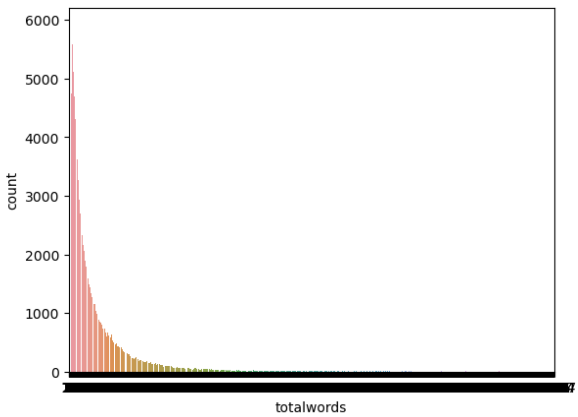
The knowledge discovered here can protect consumers and foster ethical business practices. A fundamental aspect of a well-functioning economy is access to accurate information on the part of both producers and consumers. Without accurate information, the price mechanism becomes unreliable as an economic signal and the economy becomes less efficient overall; consumers buy worse goods and services which sends the wrong signal to producers such that they start producing worse products and services. Thus everyone is harmed directly or indirectly by the prevalence of fake reviews.

The research conducted within this paper integrates machine learning and k-means models to prevent wasteful spending, enhance access to honest review platforms, and preserve the integrity of business ratings. Bridging the gap between technologies, customer well-being and businesses is something we need to have a better control on. This could even out the commerce ecosystem that has already been compromised. Our research opens up avenues for more experimentation with review datasets to further refine review evaluation models using machine learning to improve the accuracy of our prediction. more research is needed to expand on this topic with more datasets and platforms, such as amazon, ebay, walmart, etc.

Visualizations

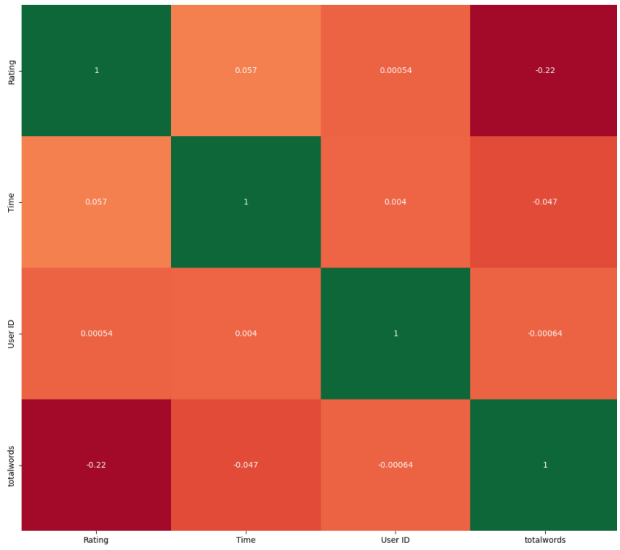
The team started the data mining efforts, focusing on analyzing reviews per day per user. The examination led to the identification of an account that raised suspicion due to having 175 reviews in a single day – an observation suggestive of potential fake reviews.

As a part of our classification analysis, we wanted to investigate word count patterns. The graph below is a count plot graph, which will aid us in establishing optimal parameters for the word count analysis.



Word Count/Review Count Graph

Subsequently, a correlation graph was generated to look at the interrelationships among our different attributes. Surprisingly, the findings lacked any strong correlations between the attributes. Instead negative correlations were found, such as “total words” and “rating”, “total words” and “time”, as well as “total words” and “user IDs”. This observation will have to be further investigated.



Attribute Correlation Heatmap

References

Clint Fontanella. 2022. How to Spot(& Remove) Fake Google Reviews. Hubspot.

DOI:https://blog.hubspot.com/service/remove-fake-google-reviews

Forhad Hossian. 2019. Fake Review Detection Using Data Mining. MSU Graduate Thesis, Missouri State University, Missouri. 1-75. DOI:https://bearworks.missouristate.edu/cgi/viewcontent.cgi?article=4454&context=theses

Myle Ott, Yejin Choi, Claire Cardie, Jeffrey Hancock. 2011. Finding deceptive opinion spam by any stretch of the imagination. In *Proceedings of the 21st international conference on World Wide Web,* pages 309-319, Association for Computational Linguistics,2011. DOI:https://aclanthology.org/P11-1032.pdf

Aishwarya Pendyala. 2019. Fake consumer review detection (Master’s thesis). California State

University, Sacramento, Department of Computer Science. 1-57. DOI:https://scholarworks.calstate.edu/downloads/xg94hv34z

Joni Salminen, Chandrashekhar Kandpal, Ahmed Mohamed Kamel, Soon-hyo Jung, Bernard J. Jansen. 2022. Creating and detecting fake reviews of online products. *Journal of Retailing and Consumer Services,* Science Direct. DOI:https://www.sciencedirect.com/science/article/pii/S0969698921003374

Aaron Smith, Monica Anderson. 2016. Pew Research Center: Online Reviews. DOI:https://www.pewresearch.org/internet/2016/12/19/online-reviews/

2019. BBC: Amazon ‘Flooded by Fake Five-Star Reviews’ - Which Report.

DOI:https://www.bbc.com/news/business-47941181