International Journal for Modern Trends in Science and Technology, 8(09): 42-48, 2022

ISSN: 2455-3778 online
DOI: https://doi.org/10.46501/IJMTST0809006
Available online at: http://www.ijmtst.com/vol8issue09.html

Fake Review Detection

Mithil Baria | Swati Mali

K J Somaiya College of Engineering, Mumbai, Maharashtra

To Cite this Article
Mithil Baria and Swati Mali. Fake Review Detection. International Journal for Modern Trends in Science and Technology 2022, 8(09), pp. 42-48. https://doi.org/10.46501/IJMTST0809006

Article Info
Received: 20 July 2022; Accepted: 26 August 2022; Published: 05 September 2022.

ABSTRACT

The world is changing to a digitized format. Users are getting more and more into digital space and have started using internet as a source of income as well as for fulfilling their necessities. The most important difference between offline market and online market, is that, in offline market a user can check how genuine a product is. But how does a user, when using online e-commerce website know the authenticity of the product. The first thing as a user, one looks for the reviews of the customers who bought the product. Now, sellers have started paying customers and different anonymous users to give good review to the product even though that product is not good. This not only tampers with the authenticity of the product but also misleads the user on purchasing the wrong product. This is where Fake Review Detection comes into picture. It is very helpful when a user comes to know how many reviews are real, helping him identify the authenticity of the product. Our project helps user identify whether the product is authentic by understanding reviews and figuring out the probability of a review being fake, using advanced machine learning algorithms.

KEYWORDS:Natural Language Processing, Reviews, Beautiful Soup, SVM, BERT, Logistic Regression.

1. INTRODUCTION

As an increasing amount of our lives is spent interacting online through social media platforms, more and more people tend to hunt out and consume news from social media instead of traditional news organizations.[11] The explanations for this alteration in consumption behaviors are inherent within the nature of those social media platforms: (i) it's often more timely and fewer expensive to consume news on social media compared with traditional journalism , like newspapers or television; and (ii) it's easier to further share, discuss , and discuss the news with friends or other readers on social media. For instance, 62 percent of U.S. adults get news on social media in 2016, while in 2012; only 49 percent reported seeing news on social media [11].

It had been also found that social media now outperforms television because the major news source. Despite the benefits provided by social media, the standard of stories on social media is less than traditional news organizations.[2] However, because it's inexpensive to supply news online and far faster and easier to propagate through social media, large volumes of faux news, i.e., those news articles with intentionally false information, are produced online for a spread of purposes, like financial and political gain. it had been estimated that over 1 million tweets are associated with fake news “Pizzagate" by the top of the presidential election. Given the prevalence of this new phenomenon, “Fake news" was even named the word of the year by the Macquarie dictionary in 2016 [12].The extensive

spread of faux news can have a significant negative impact on individuals and society. First, fake news can shatter the authenticity equilibrium of the news ecosystem for instance; it's evident that the most popular fake news was even more outspread on Facebook than the most accepted genuine mainstream news during the U.S. 2016 presidential election. Second, fake news intentionally persuades consumers to simply accept biased or false beliefs. Fake news is typically manipulated by propagandists to convey political messages or influence for instance, some report shows that Russia has created fake accounts and social bots to spread false stories.[9] Third, fake news changes the way people interpret and answer real news, for instance, some fake news was just created to trigger people's distrust and make them confused; impeding their abilities to differentiate what's true from what's not. To assist mitigate the negative effects caused by fake news (both to profit the general public and therefore the news ecosystem). It's crucial that we build up methods to automatically detect fake news broadcast on social media [3]. Internet and social media have made the access to the news information much easier and comfortable [7]. Often Internet users can pursue the events of their concern in online form, and increased number of the mobile devices makes this process even easier. But with great possibilities come great challenges. Mass media have an enormous influence on the society, and because it often happens, there's someone who wants to require advantage of this fact. Sometimes to realize some goals mass-media may manipulate the knowledge in several ways. This result in 6 producing of the news articles that isn't completely true or maybe completely false. There even exist many websites that produce fake news almost exclusively. [9]

It is now very common for people to read opinions on the web for many purposes. For example, if someone wants to buy a product and sees that the product reviews are mostly positive, they are very likely to buy the product. If the reviews are mostly negative, it is very likely that a person will choose another product. Positive opinions can lead to significant financial gains and/or fame for organizations and individuals.[3] This provides good incentives for review/opinion spam. There are generally three types of spam reviews:

Type 1 (false reviews): Those that deliberately deceive readers or opinion mining systems by giving undeservedly positive reviews to some targeted objects in order to promote those objects (which we call hyper spam) and/or by giving unfair or malicious negative reviews to some other objects to damage their reputation (what we call defamatory spam). [4]

Fake reviews are also commonly known as fake reviews or fake reviews. They have become an intense topic of discussion on blogs and forums. A recent study by Burson-Marsteller (http://www.burson-marsteller.com/Newsroom/Lists/BMNews/DispForm.aspx?ID=3645) found that an increasing number of customers are wary of fake or biased product reviews. review sites and forums. Articles about such reviews have also appeared in leading news outlets such as CNN (http://money.cnn.com/2006/05/10/news/companies/bogus_reviews/) and the New York Times (http://travel. nytimes). com/2006/02/07/business/07guides.html). These show that review spam has become a big problem. [4]

Type 2 (brand reviews only): Those who do not comment on products in product-specific reviews, but only the brands, manufacturers or sellers of the products. Although they can be useful, we consider them spam because they are not targeted at specific products and are often biased.

Type 3 (non-reviews): Those that are not reviews, which have two main subtypes: (1) advertisements and (2) other irrelevant reviews containing no opinions (eg questions, answers, and random texts).

Based on these types of spam, this article reports on a study of revision spam detection. Our research is based on 5.8 million reviews and 2.14 million reviewers (members who have written at least one review) from amazon.com. We have found that spam activities are widespread. For example, we found a large number of duplicate and near-duplicate reviews written by the same reviewers for different products, or by different reviewers (perhaps different usernames of the same people) for the same products or different products.[5]

2. RELATED WORK

Technology for Natural Language Processing (NLP) is improving day by day. Many traditional algorithms aim to provide accuracy over anything. Few papers have

also used neural networks as their primary approach for NLP. This helped them retain more information of the data's as any other. Few algorithms are mentioned below with their respective uses in NLP.

SVM

An SVM model is essentially a representation of different classes in a hyperplane in a multidimensional space. The hyperplane will be generated in an iterative way using SVM to minimize the error. The objective of SVM is to partition datasets into classes in order to find the maximum marginal hyperplane (MMH). [4]

In practice, the SVM algorithm is implemented with a kernel that transforms the input data space into the desired form. SVM uses a technique called the kernel trick, in which the kernel takes a low-dimensional input space and transforms it into a higher-dimensional space. Simply put, the kernel converts non-separable problems into separable problems by adding more dimensions. This makes SVM more powerful, flexible and accurate. [3][4]

As said above, this is a powerful algorithm, which uses hyperplane. This hyperplane will consist fake reviews on one side of the plane and other side will contain genuine reviews. Since this is an iterative approach, it provides an accurate possibility whether the review is fake or not. [4]

Logistic Regression

Logistic regression is essentially a controlled classification algorithm. In a classification problem, the target variable (or output), y, can only take on discrete values for a given set of features (or inputs), X. [5]

Contrary to popular belief, logistic regression is a regression model. The model creates a regression model to predict the probability that a given data record belongs to the category labeled "1". Just as linear regression assumes that the data follows a linear function, logistic regression models the data using a sigmoid function. [6]

Logistic regression becomes a classification technique only when a decision threshold comes into the picture. Thresholding is a very important aspect of logistic regression and depends on the classification problem itself. The decision on the threshold value is heavily influenced by the values of precision and recall. Ideally, we want both precision and recall to be 1, but that rarely happens. [5]

BERT

BERT is a transformer's pretrained model which uses attention mechanism. As RNN used neutral networks, BERT uses the same. While RNN takes each word as an input, BERT takes the whole sentence as an input. [6]

BERT uses a masked language modeling method to keep a word in focus so that it "can't see itself"—that is, it has a fixed meaning independent of its context. BERT is then forced to identify the masked word based on context alone. In BERT, words are defined by their surroundings, not by a predetermined identity. BERT is also the first NLP technique to rely solely on a self-observation mechanism, enabled by the bidirectional transformers at the heart of BERT's design. This is important because a word can often change meaning as a sentence develops. Each added word expands the overall meaning of the word targeted by the NLP algorithm. The more words in total are present in each sentence or phrase, the more ambiguous the highlighted word becomes. BERT accounts for extended meaning by reading both ways, considering the influence of all other words in the sentence on the keyword, and eliminating the left-to-right momentum that takes words toward a particular meaning over the course of a sentence. [6]

Below table is the comparison

Model Name	Strength	Weakness	Usage
SVM	Iterative approach and uses hyperplane for classification. [10]	Used on Customer Based reviews and it is difficult to maintain the same accuracy. [10]	SVM is used as an iterative approach for achieving better accuracy. [10]
Logistic Regression	Uses sigmoid function for classification. It is a regression model but can be used as classification model if	Used on Customer based dataset and does not retain the accuracy of the model. [11]	After considering all the parameters like F1-Score, Precision, Recall etc. This algorithm can be used for

	Precision and Recall parameters are considered. [11]		classification as well. [11]
BERT	Uses Attention Mechanism and is based on transformers Neural Networks. Once trained on Product based dataset, it becomes easier for text classification. [12]	Feature engineering done on dataset should be able to pass it as a parameter in BERT but is not possible since it only works on text data. [12]	Attention mechanism is used for text classification. The pretrained models are fine tuned on the datasets and then used for predictions. [12]

3. METHODOLOGY

Dataset

The dataset consists of 159402 reviews along with Reviewer name, reviewer Id, category of the products, verified purchase and the labels. This dataset was used and has been feature engineered on all the text there is. This dataset was downloaded from Kaggle (https://www.kaggle.com/datasets/naveedhn/amazon-product-review-spam-and-non-spam). In the feature engineering we found that, there are equally fake and not fake reviews. Not only this, but also verified purchasers were found to be more inclined towards genuine reviews. [13]

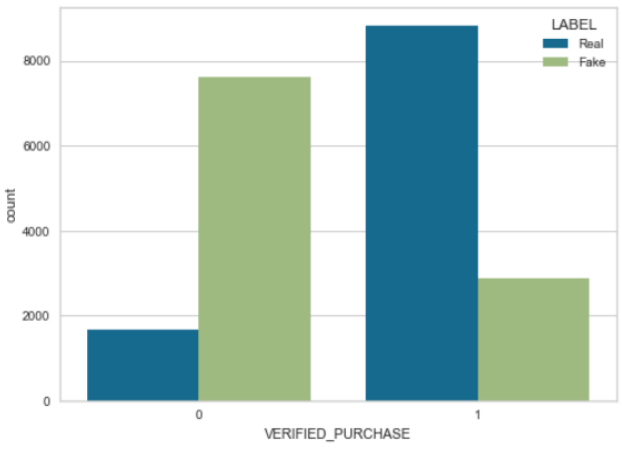

Fig 1. Verified purchase showing more inclination towards genuine reviews

The above diagram shows that most of the verified purchasers were genuine. Moreover, the length of a genuine review was larger than the fake ones. Not only the review, but also trusting the ratings was difficult because fake and non-fake reviewers almost equally gave rated the product good and bad. Below diagram shows the length of genuine review and the ratings. [13]

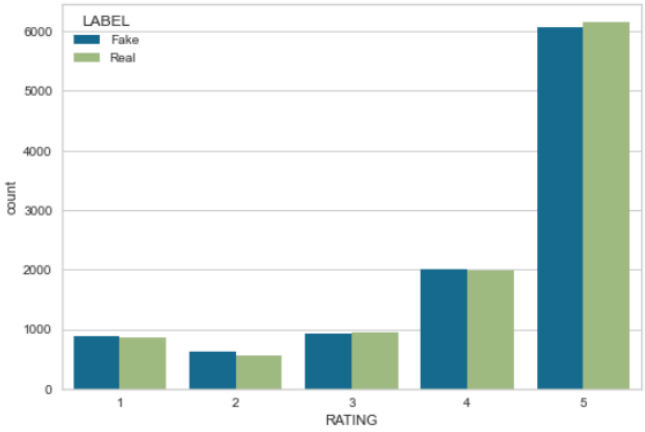

Fig 2. Ratings done equally by fake and not fake

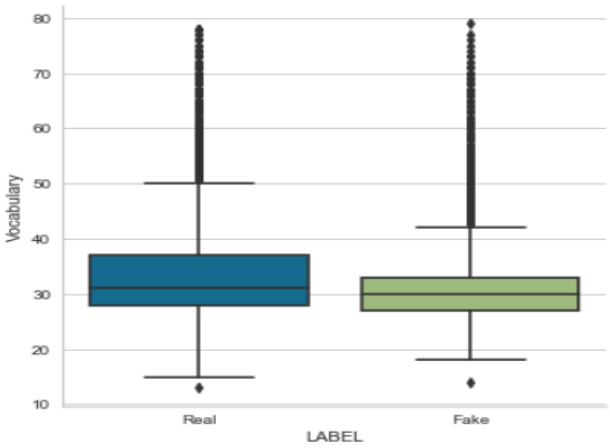

Fig 3. Length of genuine review is more

Later we have transformed the dataset into 80% - 20% split with the help of 10-Fold cross validation. [12]

The architecture consists of Data Preprocessing, Feature Engineering and Text Classification. Following shows the brief on each of them. [12]

Data Preprocessing

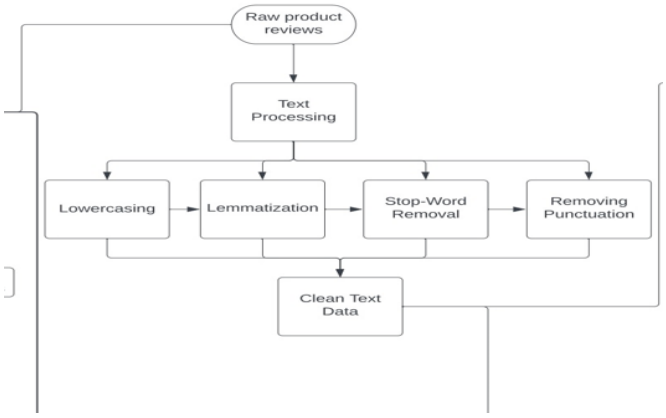

Fig 4. Data Preprocessing

In this part we pass raw data (raw reviews) for text preprocessing. This text processing consists of lowercasing, lemmatization, stop word removal and removing punctuation. Now as the name suggests, lowercasing means that we lowercase every character in the raw text. This makes easier for the model to understand the characters present in the raw data. Then comes Lemmatization, the word lemmatization means that it goes to the morphological meaning of that word which is lemma. So, if there is a sentence which consists of a word "*troubling*", it will become "*trouble*". After Lemmatization comes Stop-Word-Removal. Here we remove the words which are commonly used in a sentence. For eg. "Not, and, is, the" etc. This will give a cleaner text for the model. Later we will remove any punctuation if any, which will make the text clean enough for tokenization.

Feature Engineering

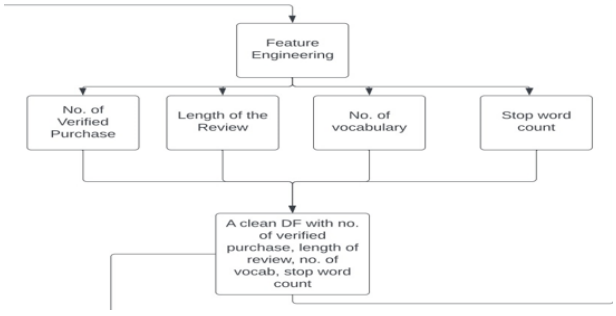

Fig 5. Feature Engineering

Feature engineering is performed to modify the data and to extract more features from the given data. So, the features like, Verified Purchaser, Length of the text, Number of words used in the text, Number of stop words present in the text and so on is included in the feature engineering. To achieve this, we have used NLTK library which has got the number of stop-words packed in it. Number of vocabulary is basically the number of words present in the corpus and all this helps us understand that how a fake review is and what a genuine review is.

Model Training

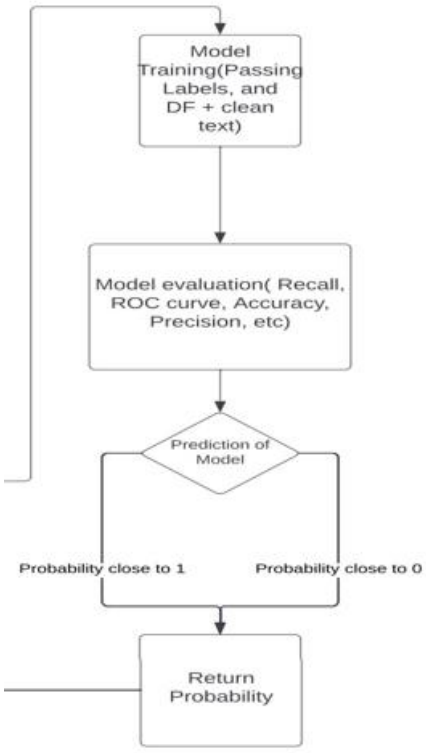

Fig 6. Model Training

Training a model is basically we pass our corpus into the model with the labels. Since our model is pre-trained, we can only fine tune it on our dataset. So changing the layers a little we can pass our dataset accordingly and train our model. Since our model is RoBerta, which is based on transformers it uses attention mechanism. RoBERTa has **a nearly** similar architecture to BERT, but to improve the results **of the** BERT architecture, the

authors made **a few** simple changes **to the** architecture and training **procedures.** These changes **include:**

Removing the Next Sentence Prediction (NSP) objective: When predicting the next **sentence,** the model is trained to predict whether the observed document segments come from the same or **different** documents **using the next sentence prediction incremental loss (NSP).** The authors **have** experimented with removing/adding NSP **losses in** different versions and **have come to the conclusion** that removing NSP **losses** matches or slightly improves **the performance of** downstream **operations. [7]**

Training with bigger batch sizes & longer sequences :Initially, BERT is trained for **1 million** steps with a batch size of 256 sequences. In this paper, the authors trained **a** model with 125 steps of **a** 2K **sequence** and 31K steps **of an** 8k batch **size sequence.** This has two **advantages. A large package reduces confusion about** the **purpose of** masked language **modeling** and **the correctness of the final work.** Large batches are also easier to parallelize **with** distributed parallel **learning. [7]**

Dynamically changing the masking pattern: In **the** BERT architecture, the masking is **done** once during data preprocessing, resulting in a single static mask. To avoid using **a** single static mask, **the** training data is **replicated** and masked 10 **times. You will have** 4 epochs with the same **mask using a different mask strategy for each 40 epochs.** This strategy is compared **to** dynamic **masking,** which **creates a** different **mask each time data** is **passed to** the model. [7]

4.RESULTS

Approach

We have performed web scrapping using python. So, when a URL is passed (Amazon URL to be precise), it will web scrape the reviews from the number of products. These reviews are then passed to the Flask app for preprocessing and then again passed in the model for prediction. The model provides us with the list of fake reviews and genuine reviews which is then calculated in percentage and displayed. The products are displayed in an ascending order by which it means, lesser the percentage genuine is the product. It is as shown in the figure below.

Results

We have considered few models such as BERT, RoBerta, XgBoost and Logistic Regression. Below table shows which model performed the best on our dataset. Here we are calculating Recall since, **Recall** is calculated as the ratio **of** the **number** of **positive** samples correctly classified as **positive** to the total number of **positive** samples. **Recall** measures the model's ability to detect positive samples. The higher the recall, the more positive samples **are found.**

Model	Recall (in percent)	F1 Score (in percent)	Accuracy (in percent)
Logistic Regression	70 – 80	65 - 73	70
XGBoost	70 – 80	70 – 80	73
BERT	75 – 83	75 – 83	64
RoBerta	92.36	92.37	82

5. FUTURE SCOPE AND CONCLUSION

To conclude, RoBerta is mostly giving accurate results when it comes to prediction or classifying reviews into fake and genuine. Also, since RoBerta takes only text as an input, the other features are getting extracted while training, but a model in which we can also pass these attributes can make it easier for training and prediction. Moreover, we can make this as a plugin which will do this work just by a single click.

Conflict of interest statement

Authors declare that they do not have any conflict of interest.

REFERENCES

[1] https://huggingface.co/docs/transformers/model_doc/bert

[2] https://huggingface.co/transformers/v3.0.2/model_doc/bert.html

[3] https://www.tutorialspoint.com/machine_learning_with_python/machine_learning_with_python_classification_algorithms_support_vector_machine.htm#:~:text=Working%20of%20SVM,maximum%20marginal%20hyperplane%20(MMH).

[4] https://www.geeksforgeeks.org/understanding-logistic-regression/
[5] https://www.analyticsvidhya.com/blog/2021/07/an-introduction-to-logistic-regression/#:~:text=Logistic%20Regression%20is%20a%20%E2%80%9CSupervised,used%20for%20Binary%20classification%20problems.
[6] https://www.geeksforgeeks.org/overview-of-roberta-model/
[7] https://ai.facebook.com/blog/roberta-an-optimized-method-for-pretraining-self-supervised-nlp-systems/
[8] Ata-Ur-Rehman et al., "Intelligent Interface for Fake Product Review Monitoring and Removal," 2019 16th International Conference on Electrical Engineering, Computing Science and Automatic Control (CCE), 2019, pp. 1-6, doi: 10.1109/ICEEE.2019.8884529.
[9] https://dl.acm.org/doi/10.1145/3419604.3419800
[10] R. Mohawesh et al., "Fake Reviews Detection: A Survey," in IEEE Access, vol. 9, pp. 65771-65802, 2021, doi: 10.1109/ACCESS.2021.3075573.
[11] R. Hassan and M. R. Islam, "A Supervised Machine Learning Approach to Detect Fake Online Reviews," 2020 23rd International Conference on Computer and Information Technology (ICCIT), 2020, pp. 1-6, doi: 10.1109/ICCIT51783.2020.9392727.
[12] https://www.cs.uic.edu/~liub/FBS/opinion-spam-WSDM-08.pdf
[13] https://www.kaggle.com/datasets/naveedhn/amazon-product-review-spam-and-non-spam