Semi-Supervised Learning based Fake Review Detection

Sushama Sharad Khilari[1]
Department of Information Technology
Pillai College of Engineering
(University of Mumbai)
Mumbai, India

Kalpesh Ashok Patil[2]
Department of Information Technology
Pillai College of Engineering
(University of Mumbai)
Mumbai, India

Kiran Krishna Solunke[3]
Department of Information Technology
Pillai College of Engineering
(University of Mumbai)
Mumbai, India

Aju palleri[4]
Department of Computer Engineering
Pillai College of Engineering
(University of Mumbai)
Mumbai, India

***Abstract*—** The use of Internet and online marketing has become immensely popular. Countless number of products and services are available in online markets. Therefore, it is challenging enough to find the best suitable services or products compatible to our requirement. Customers directly assume a review or opinion written by others without second thought. Any person can write anything as a review which is why we have to deal with this problem of fake reviews. Various companies are hiring people to write fake positive reviews about their services or products or unfair negative reviews to their competitors' services or products. Due to this the new customers who wish to buy such items are left with misleading or ingenuine reviews and hence we need a system to detect them and remove them. In this paper we discuss various supervised data mining techniques for fake review detection.

I. INTRODUCTION

The traditional algorithms rely on human developers to provide explicit instructions, machine learning allows data to guide the process. Using a training set comprised of representative data, the computer develops a model that can be compared against an expected output. The more data the computer takes into account, the more accurate and sophisticated its model becomes without developers having to write additional lines of code. The same advancements in storage and processing that have facilitated the rise of big data have also allowed machine learning to advance by leaps and bounds. Storage and processing frameworks like Hadoop and robust libraries for programming languages like R, Python, and Java have made it feasible for the first time to feed machine learning algorithms tremendous amounts of unstructured data. For data scientists, the ability to create algorithms that are guided by actual data is a huge advantage. For data scientists, the ability to create algorithms that are guided by actual data is a huge advantage.

A. Objectives

a) To build a model that can differentiate between 'real' and 'fake' news.
b) To design and implement a system that overcomes the drawbacks of the existing systems.
c) To design a system that is Cost effective, reliable and consumes less power.

B. Scope

Machine Learning is currently one of the hottest topics in IT. Technologies such as digital, big data, Artificial Intelligence, automation and machine learning are increasingly shaping future of work and jobs. is a specific set of techniques that enable machines to learn from data, and make predictions. When the biases of our past and present fuel the predictions of the future, it's a tall order to expect AI to operate independently of human flaws. The goal of our project is to create a model which can be used in social platform to detect any fake news which user shared.

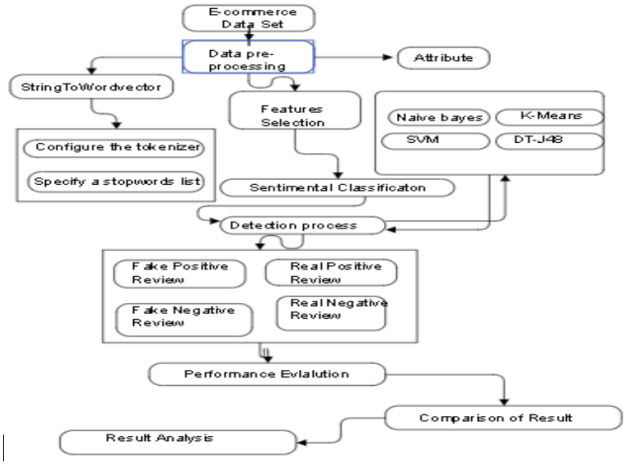
II. SYSTEM ARCHITECTURE

A. Existing System

Existing system they proposed several methods to analyze a dataset of movie reviews. They also presented sentiment classification algorithms to apply a supervised learning of the movie reviews located in two different datasets. The experimental approaches studied the accuracy of all sentiment classification algorithms, and how to determine which algorithm is more accurate. Furthermore, we were able to detect fake positive reviews and fake negative reviews through detection processes. Two supervised learning algorithms to classifying sentiment of dataset: NB and K*. Using the accuracy analysis for these two techniques

Zheg-Jun-Zha first identify product aspects by a shallow dependency parser and determine consumer opinions on these aspects via a sentiment classifier. Then develop a probabilistic aspect ranking algorithm to infer the importance of aspects by simultaneously considering aspect frequency and the influence of consumer opinions given to each aspect over their overall opinions. The experimental results on a review corpus of 21 popular products in eight domains demonstrate the effectiveness of the proposed approach. Moreover, we apply product aspect ranking to two real-world applications, i.e., document-level sentiment classification and extractive review summarization, and achieve significant performance improvements, which demonstrate the capacity of product

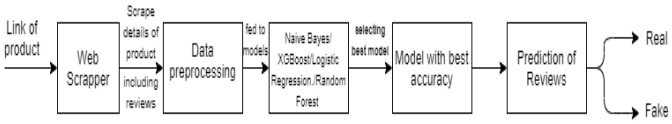
Existing system block diagram



B. Proposed System

In this approach we build a model based on the count vectorizer or a tfidf matrix (i.e) word tallies relatives to how often they are used in other artices in your dataset. Since this problem is a kind of text classification, Implementing a Naive Bayes classifier will be best as this is standard for text-based processing. However, we also test other machine learning models like XgBoost, Random Forest and so on. Now the next step is to extract the most optimal features for countvectorizer or tfidf-vectorizer, this is done by using a n-number of the most used words, and/or phrases, lower casing or not, mainly removing the stop words which are common words such as “the”, “when”, and “there” and only using those words that appear at least a given number of times in a given text dataset.

C. System Block Diagram



D. Implementation Details

The implementation details are given in this section.

Module 1 - Data Collection

The dataset is the primary requirement in this problem. For our implementation we use the product reviews on amazon using web scraping.

Module 2 - Data Pre-Processing

The preprocessing phase includes two preliminary operations, shown in Figure 1, that help in transforming the data before the actual SA task. Data preprocessing plays a significant role in many supervised learning algorithms. We divided data preprocessing as follows:

1) Generating Vectors for the text data

Machine learning models cannot be operated on text data. Hence, we need to convert our text data into numerical form before feeding it to machine learning models. Our text data will be converted into vectors using TFIDF and Bag of Words (BOW) algorithms.

2) Attribute Selection

Removing the poorly describing attributes can significantly increase the classification accuracy, in order to maintain a better classification accuracy, because not all attributes are relevant to the classification work, and the irrelevant attributes can decrease the performance of the used analysis algorithms, an attribute selection scheme was used for training the classifier.

3) Stopword Removal: The stopwords are the words we want to filter out, eliminate, before training the classifier. Some of those words are commonly used (e.g., "a," "the," "of," "I," "you," "it," "and") but do not give any substantial information to our labeling scheme, but instead they introduce confusion to our classifier.

Module 3 - : Feature Selection

Feature selection is an approach which is used to identify a subset of features which are mostly related to the target model, and the goal of feature selection is to increase the level of accuracy. In this study, we implemented four feature selection methods widely used for the classification task of SA with Stopwords methods.

Module 4 Detection Processes

After training, the next step is to predict the output of the model on the testing dataset, and then a confusion matrix is generated which classifies the reviews as positive or negative. The results involve the following attributes:

True Positive: Real Positive Reviews in the testing data, which are correctly classified by the model as Positive (P).

False Positive: Fake Positive Reviews in the testing data, which are incorrectly classified by the model as Positive (P).

True Negative: Real Negative Reviews in the testing data, which are correctly classified by the model as Negative (N).

False Negative: Fake Negative Reviews in the testing data, which are incorrectly classified by the model as Negative (N).

True negative (TN) are events which are real and are effectively labeled as real, True Positive (TP) are events

which are fake and are effectively labeled as fake. Respectively, False Positives (FP) refer to Real events being classified as fakes;

False Negatives (FN) are fake events incorrectly classified as Real events.

Predictions obtained with known data, and for each algorithm used in this.

Naive Bayes Algorithm

Naive Bayes is a simple technique for constructing classifiers: models that assign class labels to problem instances, represented as vectors of feature values, where the class labels are drawn from some finite set[14].Abstractly Naive Bayes is a conditional probabilistic model: given a problem instance to be classified, represented by a vector x=(x1,x2,....xn) representing some *n* features (independent variables), it assigns to thisinstance probabilities.

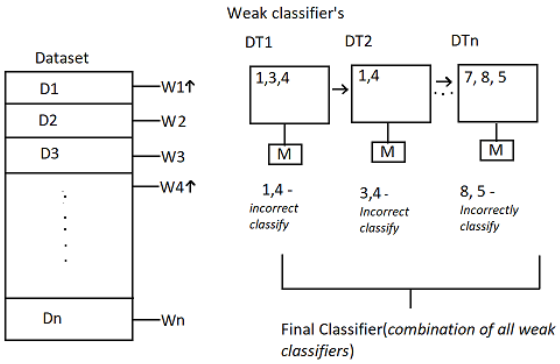
$$p(C_k \mid x_1, \ldots, x_n)$$

In plain English, using Bayesian probability terminology, the above equation can be written as

$$\text{posterior} = \frac{\text{prior} \times \text{likelihood}}{\text{evidence}}$$

Xgboost algorithm

It is collection of decision trees. Initially the dataset assign same weights. Each decision tree contain model which will do predictions. It is possible that model do misclassifications(weak classifiers). Update(increase) weights. Then next DT will take max weights and do prediction. Again - misclassifications(weak classifiers). Update(increase) weights. And finally it take majority.



Random Forest

Random forests or random decision forests are an ensemble learning method for classification, regression and other tasks that operate by constructing a multitude of decision trees at training time and outputting the class that is the mode of the classes (classification) or mean prediction (regression) of the individual trees. Random decision forests correct for decision trees' habit of overfitting to their training set.

Logistic Regression

In statistics, the logistic model (or logit model) is used to model the probability of a certain class or event existing such as pass/fail, win/lose, alive/dead or healthy/sick. This can be extended to model several classes of events such as determining whether an image contains a cat, dog, lion, etc. Each object being detected in the image would be assigned a probability between 0 and 1 and the sum adding to one.

III. HARDWARE SPECIFICATION

Processor	Intel(R) Core(TM) i7-8550U CPU @ 1.80GHz, 1992 Mhz, 4 Core(s), 8 Logical Processor(s)
OS Name	Microsoft Windows 10 Home Single Language
RAM	8 GB DDR4
Internal Memory	2TB

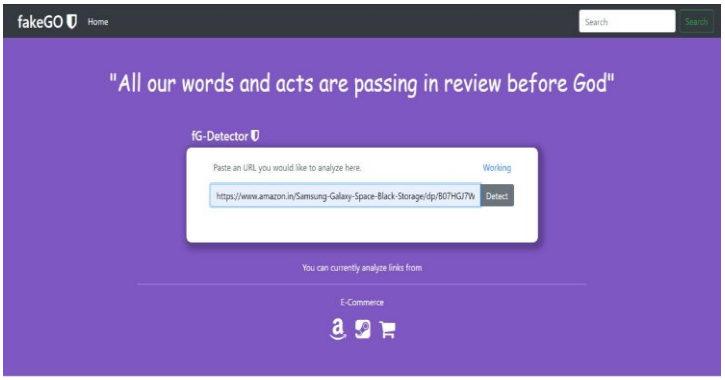
IV. SOFTWARE SPECIFICATION

Jupyter notebook
Sublime Text IDE

V. PROJECT INPUTS AND OUTPUTS

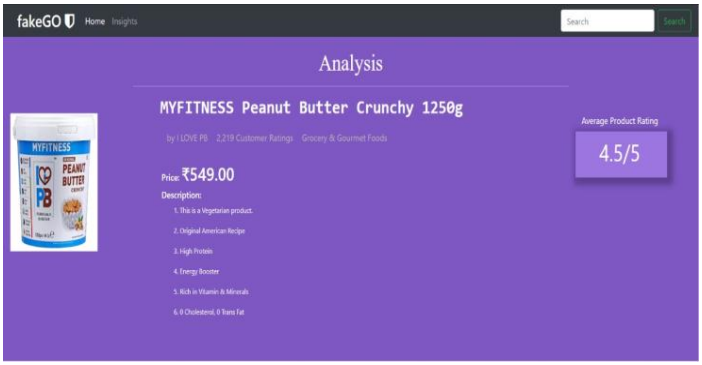
A. Input details.

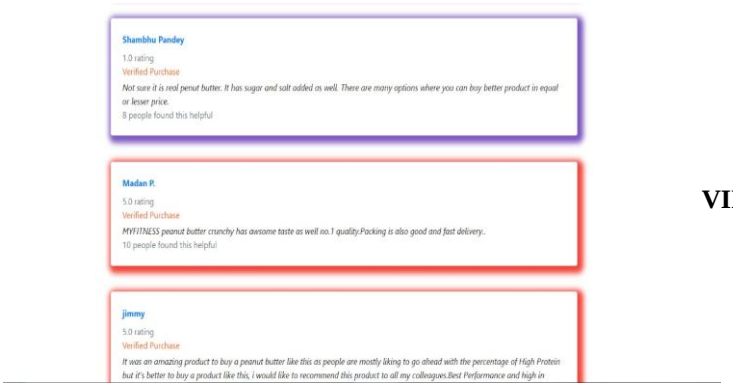
User can enter product link from amazon.in on the below textbox and our algorithm will use a web crawler to scrape all the reviews and highlight all the fake reviews. User will also be able to analyze the results from different models using a pie chart.

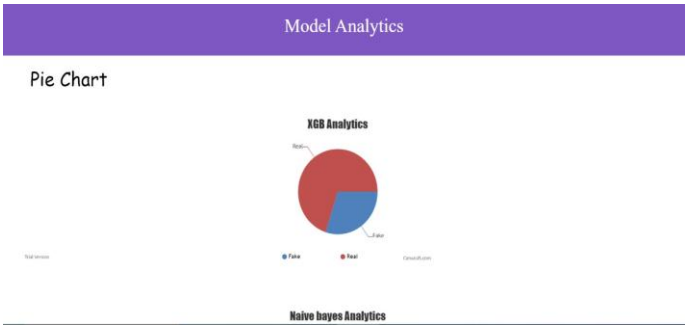


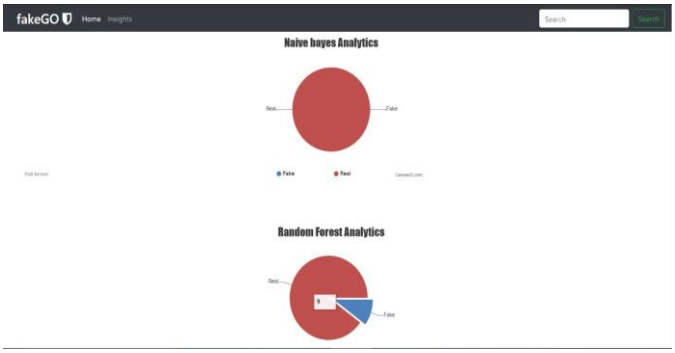
B. Output Details

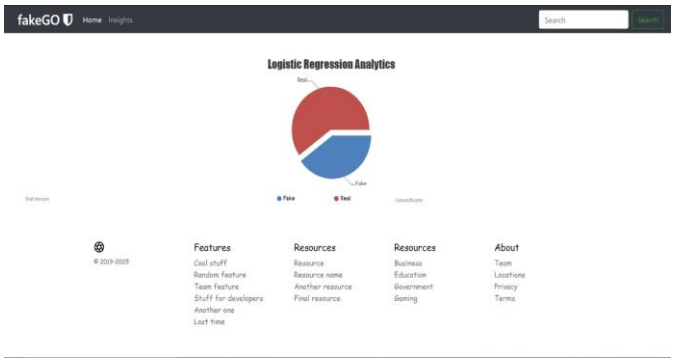
Following is the sample output which will be displayed when user enters a product link.











VI. APPLICATIONS

- In e-commerce websites, it can be used to filter out all the ingenuine reviews.
- It could also be used as a mobile application to verify if a review is fake or genuine.
- Anti-spam system used to assess which review is genuine, and which is fake.

VII. SUMMARY

In this paper, we explored the fake reviews problem by reviewing existing literature and identified the flaws in them and have built a very comprehensive, simple yet an effective solution for the same. We also further discussed the datasets, evaluation metrics, and promising future directions in fake review classification research and expand the field to other applications.

REFERENCES

[1]Semi-supervised Learning based Fake Review Detection(2017), Huaxun Deng, Linfeng Zhao.

[2]Detecting Fake Reviews through Sentiment Analysis Using Machine Learning Techniques Elshrif Elmurngi, Abdelouahed Gherbi

[3]Semi-supervised Learning based Spammer Group Detection in Product Reviews(2017) Lu Zhang†, Yang Yuan

[4]Fake Review Detection from A Product Review Using Modified Method Of Iterative Computation Framework, Eka Dyar Wahyuni And ARIF Junaidi

[5]Detection of fake online hotel reviews (2017) Casey Wilson ,Anna V. Sandifer

[6]Fake Review Detection: Classification and Analysis of Real and Pseudo Reviews Arjun Mukherjee†, Vivek Venkataraman

[7]Fraud Detection in Online Reviews using Machine Learning Techniques ,Kolli Shivagangadhar, Sagar H, Sohan'

[8]Detection of Fake Review and Brand Spam Using Data Mining Technique (2016),Miss.Rashmi Gomatesh Adike1 , Prof. Vivekanand Reddy2

[9]Reliable Fake Review Detection via Modeling Temporal and Behavioral Patterns, Xian Wu, Yuxiao Dong, Jun Tao, Chao Huang, Nitesh V. Chawla

[10]Detecting Singleton Review Spammers Using Semantic Similarity ,Vlad Sandulescu Martin Ester

[11]Methods And Approches On Spam Review Detection For Sentimental Analysis ,Sindhu C, G. Vadivu, Anirudh Singh, Rahil Patel

[12]Sentiment Analysis Based Online Restaurants Fake Reviews Hype Detection, Xiaolong Deng 1 and Runyu Chen 2