International Journal of Scientific Research in Computer Science, Engineering and Information Technology
ISSN : 2456-3307 (www.ijsrcseit.com)
doi : https://doi.org/10.32628/IJSRCSEIT

Identification of Fake Reviews Using Supervised Machine Learning

M. Prathap Reddy[1] G.Lakshmikanth[2]
M Tech Student[1], Associate Professor[2]
Department of Computer Science and Engineering, Sree Rama Engineering College, Tirupati, Andhra Pradesh, India

Article Info

Publication Issue :
Volume 8, Issue 6
November-December-2022

Page Number : 116-124

Article History
Accepted: 05 Nov 2022
Published: 15 Nov 2022

ABSTRACT

Online reviews are largely regarded as a significant aspect for establishing and preserving a solid reputation as e-commerce systems continue to advance. Additionally, they play a significant part in how end customers decide. A favorable review for a specific item typically draws in more customers and increases sales significantly. In order to develop a virtual reputation and draw in new clients, reviews that are false or misleading are being intentionally written. Therefore, spotting bogus reviews is an active and developing study field. The ability to spot false reviews depends on both the essential characteristics of the reviews and the behaviour of the reviewers. This study suggests using machine learning to spot bogus reviews. In addition to the features extraction process of the reviews, this paper applies several features engineering to extract various behaviours of the reviewers. The performance of many experiments conducted on a real dataset of restaurant reviews from Yelp is compared in this research, including KNN, Naive Bayes (NB), and Logistic Regression. The findings show that Logistic Regression performs better than the other classifiers in terms of accuracy. The findings demonstrate that the algorithm is better able to distinguish between authentic and false reviews.

Keywords : Machine learning, fake, reviews, Logistic Regression.

I. INTRODUCTION

Reviews have replaced other sources of information for consumers looking to make decisions regarding services or items in the modern day. For instance, when clients decide to book a hotel, they read reviews about what other guests have to say about the hotel services. They determine whether or not to make a reservation based on the feedback from the reviews. If they found the evaluations to be favorable, they will probably go ahead and reserve the room. As a result, historical analyses rose to the top of many web services' lists of highly reliable information sources. Reviews are seen as real feedback regarding good or bad services, therefore any attempt to skew them by including false or misleading information is seen as dishonest behaviour and is accompanied by the label "fake reviews." Such a situation makes us wonder what would happen if not all submitted reviews are trustworthy or sincere. What if any of these testimonials are false? As a result, the detection of fraudulent reviews has been and is currently an active and important study subject.

The emergence of social media has made it more difficult to distinguish between genuine material and advertising, which has caused an increase of false endorsements throughout the industry. Online advertisements for items frequently use false testimonials and other misleading endorsements. In order to remind marketers of the law and dissuade them from breaching it, the FTC is now employing its penalty offence authority. The agency is warning more than 700 businesses that using endorsements in ways that are inconsistent with earlier FTC administrative proceedings might result in hefty civil fines (up to $43,792 per violation).

- According to Samuel Levine, director of the Federal Trade Commission's Bureau of Consumer Protection, "Fake reviews and other types of fraudulent endorsements defraud customers and undermine honest businesses." If advertisers use these dishonest tactics, they will pay a price.
- The Notice of Penalty Offenses empowers the FTC to pursue civil penalties against a business whose actions it is aware have been deemed illegal in a prior administrative order, other than a consent order, by informing the corporation that its actions have been ruled unlawful..

The FTC identified a number of behaviours in the Notice that it ruled to be unfair or misleading in earlier administrative actions. The Notice was addressed to the firms. These include, but are not limited to: falsely claiming an endorsement by a third party; misrepresenting that an endorser is an actual, current, or recent user; using an endorsement to make false performance claims; failing to disclose an unexpected material connection with an endorser; and falsely representing that an endorser's experience represents consumers' typical or ordinary experience.

The recipients of the notification include a wide range of big businesses, eminent advertisers, prominent retailers, top manufacturers of consumer goods, and significant advertising agencies. The FTC's website has a complete list of the companies that have received the Notice. The inclusion of a recipient on this list in no way implies that it has participated in dishonest or unfair behaviour.

In addition to the Notice, the FTC has developed a number of resources that can be available on the FTC website to help businesses make sure they are abiding with the law when utilizing endorsements to market their goods and services.

In order to do this, this study utilizes a number of machine learning classifiers to spot false reviews based on the reviewers' own attributes as well as the content of the reviews. On a genuine corpus of reviews obtained from open source websites, we apply the classifiers. The research utilizes a number of features engineering techniques to the corpus in addition to the standard natural language processing to extract and feed the characteristics of the reviews to the classifiers. The research examines the effects of reviewers' extracted characteristics when they are taken into account by the classifiers. The study examines the outcomes of two alternative language models, TF-IDF and those without the extracted characteristics. The findings show that the built features improve the effectiveness of the method for identifying bogus reviews.

Many of our everyday activities have been impacted by the Internet's explosive rise. E-commerce is one of the sectors with the fastest development. The majority of e-commerce sites allow users to post evaluations of their services. These reviews' presence can be used as a source of data. For instance, businesses may use it to select how to design their goods or services, and prospective customers can use it to choose whether to purchase or use a product. Unfortunately, some people have tried to manufacture phone reviews in an effort to either boost the popularity of the product or discredit it, taking advantage of its significance. This study uses the language and rating properties from a review to identify phone product reviews. Many of our everyday activities have been impacted by the Internet's explosive rise. Ecommerce is one of the sectors with the fastest development. Typically, online stores allow customers to post evaluations of their services. These reviews' presence can be used as a source of knowledge. For instance, businesses may use it to select how to design their goods or services, and prospective customers can use it to choose whether to purchase or use a product. Unfortunately, some people have tried to manufacture phone reviews in an effort to either boost the popularity of the product or discredit it, taking advantage of its significance. This study uses a review's language and rating information to identify phone product reviews..

Many of our everyday activities have been impacted by the Internet's explosive rise. E-commerce is one of the sectors with the fastest development. The majority of e-commerce sites allow users to post evaluations of their services. These reviews' presence can be used as a source of knowledge. For instance, businesses may use it to select how to design their goods or services, and prospective customers can use it to choose whether to purchase or use a product. Unfortunately, some people have tried to manufacture phone reviews in an effort to either boost the popularity of the product or discredit it, taking advantage of its significance. The goal of this study is to identify false product reviews using the text and rating information from a review. The detection of false evaluations of web materials may greatly benefit from machine learning approaches. In general, web mining approaches utilize a variety of machine learning algorithms to identify and extract important information. Content mining is one of the web mining duties. Opinion mining is a classic illustration of content mining since it uses machine learning to identify the sentiment of text (positive or negative), and a classifier is developed to analyse both the characteristics and the sentiment of reviews. The identification of phone reviews often focuses on variables that are not directly related to the content as well as the category of the reviews. Text and natural language processing are typically used in developing review feature sets NLP. However, including factors relating to the reviewer himself, such as the review time/date or his writing styles, may be necessary to create phone reviews. Therefore, the development of useful characteristics extraction from the reviewers is key to the effective identification of false reviews.

The identification of phone reviews often focuses on variables that are not directly related to the content as well as the category of the reviews. Text and natural language processing NLP are typically used while creating review feature sets. However, including factors relating to the reviewer himself, such as the review time/date or his writing styles, may be necessary to create phone reviews. Therefore, the development of useful characteristics extraction from the reviewers is key to the effective identification of false reviews.

II. RELATED WORKS

A framework for fake review detection in online consumer electronics retailers: Online evaluations now have a huge influence on organizations, and they are essential to determining business performance in a variety of industries, from hotels and restaurants to e-commerce. Unfortunately, some individuals write fictitious evaluations of their own companies or rivals in an effort to boost their online image. Previous studies have focused on the identification of fraudulent reviews in a variety of fields, including restaurant and hotel product or company reviews. The area of consumer electronics enterprises has not yet been adequately explored, despite its economic relevance. This article suggests a feature structure that has been tested in the consumer electronics industry for identifying false reviews. There are four contributions: Building a dataset for categorizing fake reviews in the consumer electronics sector using scraping techniques in four different cities; defining a feature framework for fake review detection; developing a method for categorizing fake reviews based on the proposed framework; and evaluating and analysing the results for each of the cities under study. The Ada Boost classifier has been shown to be the best by statistical methods according to the Friedman test, and we have achieved an 82% F-Score on the classification challenge.

The economics of reputation and feedback systems in e-commerce marketplaces: Online markets are already commonplace because to the massive user bases of websites like eBay, Taobao, Uber, and Airbnb. These marketplaces' success is due in part to how simple it is for buyers to locate sellers as well as to the trust that is fostered by their reputation and feedback systems. I start by providing an overview of feedback and reputation systems used in online markets, as well as a quick description of the fundamental concepts around the function of reputation in fostering trust and commerce. The research that investigates the impact of reputation and feedback systems on online markets is then described, and some of the bias issues with current feedback and reputation systems are highlighted. I present solutions to these issues to enhance the usability of online marketplace designs and offer some suggestions for further study.

Data mining: Web data mining techniques, tools and algorithms: An overview: Web data mining has developed into a simple and crucial platform for the retrieval of pertinent information. Users choose to upload and download material over the World Wide Web. Finding useful information and patterns is becoming more difficult and time-consuming as the amount of data on the internet grows. It is difficult to extract accurate, user-requested information from unstructured, inconsistent material on the internet. To retrieve pertinent information from the web, many mining techniques are utilized (hyperlinks, contents, web usage logs). A kind of data mining called "web data mining" focuses mostly on the web. Web use, web content, and web structure mining are the three different categories of web data mining. All of these categories employ various methods, devices, strategies, and algorithms to extract information from vast quantities of online data.

Opinion mining and sentiment analysis: Users may now share opinions about things, people, events, and subjects in a variety of official and informal venues thanks to the recent growth of social media. These options include blogs, discussion boards, social media postings, forums, and reviews. The computational analytics relating to such text are referred to as the opinion mining and sentiment analysis problem.

What yelp fake review filter might be doing: Online reviews have developed into a useful tool for selecting choices. However, the benefit of it is accompanied by a curse: false opinion spam. The identification of false reviews has received a lot of attention recently. The majority of review websites do not, however, publicly screen out fraudulent reviews. One exception is Yelp, which has been removing reviews for the past five years. Yelp's algorithm is a trade secret, though. In this study, we analyse Yelp's filtered reviews in an effort to understand what the company may be up to. Other review hosting websites will find the results helpful in their screening efforts. supervised and unsupervised learning are the two basic methods of filtering. There are also essentially two sorts of characteristics that are used: Language and behavioural characteristics. We'll utilize a supervised strategy in this project because we can train using Yelp's filtered reviews. All currently used supervised learning methods are based on pseudo-fake reviews rather than false reviews that have been vetted by a for-profit website. Recently, it has been demonstrated that supervised learning utilizing linguistic n-gram features works remarkably well (with over 90% accuracy) in identifying crowdsourcing fraudulent reviews made via Amazon Mechanical Turk (AMT). We examine these currently used research methodologies and assess their effectiveness using actual Yelp data. To our surprise, the language characteristics do not perform as well as the behavioural aspects. A fresh information theoretic methodology is suggested to study and determine the precise psycholinguistic distinction between Yelp and AMT reviews (crowd sourced vs. commercial fake reviews). We come upon something quite intriguing. We may assume that Yelp's filtering is appropriate and that its filtering method appears to be connected with anomalous spamming activities thanks to this study and testing findings.

Review spam detection: Today, it is standard practice for e-commerce websites to offer the option for users to leave reviews for the goods they have purchased. These reviews serve as useful resources for knowledge about these goods. They are used by prospective buyers to research user reviews before making a purchase. Additionally, they help product makers detect issues with their goods and gather competitive knowledge about their rivals. Sadly, the value of reviews serves as a strong motivator for spam that incorporates deceptive positive or maliciously negative remarks. We aim to examine review spam and spam detection in this essay. To the best of our knowledge, no known study has been done on this issue.

III. METHODOLOGY

Proposed system:

We suggest this application, which may be seen as a valuable system since it aids in reducing the constraints brought about by conventional and other existing ways. The goal of this study is to provide a quick, dependable approach for correctly detecting and estimating anemia. We employed a sophisticated algorithm to develop this system in a Python environment using the Django framework.

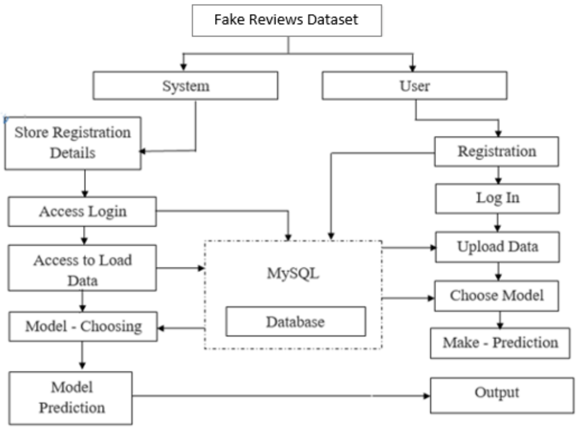


Figure 1 : Block diagram

IV. IMPLEMENTATION

The project has implemented by using below listed algorithm.

1. Naive Bayes:

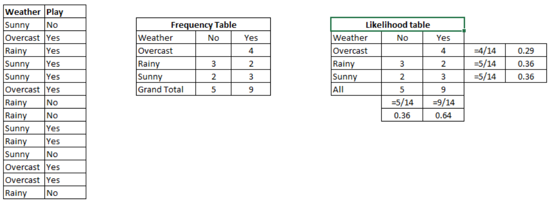
A probabilistic machine learning model called a Naive Bayes classifier is utilized for classification tasks. The Bayes theorem is the cornerstone of the classifier.

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

When B has already happened, we may use the Bayes theorem to calculate the likelihood that A will also occur. Here, A is the hypothesis and B is the supporting evidence. Here, it is assumed that the predictors and characteristics are independent. That is, the existence of one characteristic does not change the behaviour of another. The term "naive" is a result. Let's use an illustration to comprehend it. I've included a training set of weather data and its matching goal variable, "Play," below (suggesting possibilities of playing). We must now categories whether participants will participate in games based on the weather. Let's carry it out by following the instructions below.

Step 1: Convert the data collection into a frequency table.

Step 2: Make a likelihood table by calculating probabilities such as Overcast probability = 0.29 and Playing probability = 0.64.



Weather	Play
Sunny	No
Overcast	Yes
Rainy	Yes
Sunny	Yes
Sunny	Yes
Overcast	Yes
Rainy	No
Rainy	No
Sunny	Yes
Rainy	Yes
Sunny	No
Overcast	Yes
Overcast	Yes
Rainy	No

Frequency Table		
Weather	No	Yes
Overcast		4
Rainy	3	2
Sunny	2	3
Grand Total	5	9

Likelihood table				
Weather	No	Yes		
Overcast		4	=4/14	0.29
Rainy	3	2	=5/14	0.36
Sunny	2	3	=5/14	0.36
All	5	9		
	=5/14	=9/14		
	0.36	0.64		

Step 3: In order to determine the posterior probability for each class, utilize the Naive Bayesian equation. The result of a prediction is the class with the highest posterior probability.

Problem: In bright conditions, players will still show up. Does this claim hold true?

We can resolve it utilizing the posterior probability approach that was just presented.

P (Yes | Sunny) = P(Sunny | Yes) * P(Yes) / P (Sunny)

Here we have P (Sunny |Yes) = 3/9 = 0.33, P (Sunny) = 5/14 = 0.36, P (Yes) = 9/14 = 0.64

Now, P (Yes | Sunny) = 0.33 * 0.64 / 0.36 = 0.60, which has higher probability.

A similar approach is used by Naive Bayes to forecast the likelihood of various classes based on various characteristics. The main use of this approach is text classification, which has issues with having several classes.

• Class of test data set prediction is quick and simple. Additionally, it excels in multi-class prediction.

• A Naive Bayes classifier performs better when the assumption of independence is true than other models, such as logistic regression, and requires less training data.

• When compared to numerical input variables, it performs well with categorical input variables (s). It is assumed that numerical variables have a normal distribution (bell curve, which is a strong assumption).

Applications of Naive Bayes Algorithms

- Real time Prediction
- Multi class Prediction
- Text classification/ Spam Filtering/ Sentiment Analysis
- Recommendation System

KNN:

- One of the simplest machine learning algorithms, based on the supervised learning method, is K-Nearest Neighbours.
- The K-NN method places the new case in the category that is most comparable to the available categories, assuming that the new case/data and the existing cases are similar.
- The K-NN algorithm saves all of the information that is available and categorizes new data based on similarity. This means that utilizing the K-NN method, fresh data may be quickly and accurately sorted into a suitable category.
- The K-NN approach may be used for both classification and regression, however it is most frequently utilized for classification tasks.
- K-NN is a non-parametric method, which means it makes no assumptions about the underlying data.
- It is also known as a lazy learner algorithm since it doesn't immediately apply what it has learned to the training set; instead, it saves the information and applies it to the classification process.

The KNN method simply saves the information during the training phase, and when it receives new data, it categorizes it into a category that is quite similar to the new data.

The following algorithm may be used to demonstrate how K-NN works:

- ○ **Step-1:** Select the number K of the neighbors
- ○ **Step-2:** Calculate the Euclidean distance of K number of neighbors
- ○ **Step-3:** Take the K nearest neighbors as per the calculated Euclidean distance.

- **Step-4:** Among these k neighbors, count the number of the data points in each category.
- **Step-5:** Assign the new data points to that category for which the number of the neighbor is maximum.
- **Step-6:** Our model is ready.

3. Logistic Regression:

Early in the 20th century, the biological sciences began to employ logistic regression. Then, it was put to many different social science uses. When the dependent variable (target) is categorical, logistic regression is utilized.

For instance,

Determining whether an email is spam (0)

Whether or if the tumor is malignant (0)

Consider a situation where we must determine whether or not an email is spam. In order to do classification if we utilize linear regression to solve this issue, a threshold has to be established. Say the data point is classed as nonmalignant even though the actual class is malignant, anticipated continuous value is 0.4, and the threshold value is 0.5. This might have catastrophic consequences in the present.

From this illustration, it can be concluded that classification problems do not lend themselves to linear regression. Logistic regression enters the scene as a result of the unlimited nature of linear regression. They only have values between 0 and 1.

Uses of logistic regression:

Online advertising has benefited greatly from the growing popularity of logistic regression since it allows advertisers to forecast the likelihood, expressed as a percentage, of individual website visitors clicking on particular adverts.

- In order to determine disease risk variables and develop preventative strategies, healthcare facilities can also employ logistic regression.
- Voting applications to anticipate if voters would support a certain candidate; • Weather forecasting apps to forecast snowfall and weather conditions.

Insurance that estimates the likelihood that a policyholder would pass away before the policy's term expires using certain factors, such as gender, age, and physical examination.

Using yearly income, prior defaults, and historical debts, banking can forecast whether a loan application would default or not.

XGBoost:

XGBoost stands for "Extreme Gradient Boosting" is what the acronym. XGBoost is a distributed gradient boosting library that has been developed to be very effective, adaptable, and portable. It uses the Gradient Boosting framework to construct machine learning algorithms. It offers a parallel tree boosting to quickly and accurately address a variety of data science challenges.

Boosting

Boosting is an ensemble learning strategy that creates a strong classifier out of a number of successively weak classifiers. In order to address the bias-variance trade-off, boosting techniques are essential. Boosting algorithms are thought to be more successful than bagging algorithms, which merely take into account the large variation in a model.

Below are the few types of boosting algorithms:

- AdaBoost (Adaptive Boosting)
- Gradient Boosting
- XGBoost
- CatBoost
- Light GBM

XGBoost

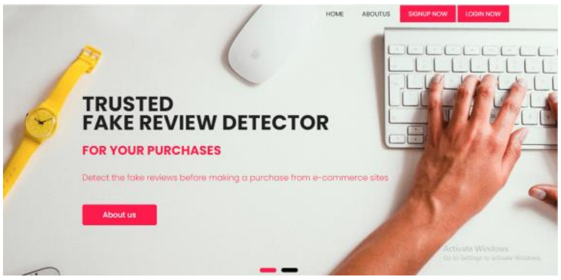
Excessive gradient boosting is referred to as XGBoost. Due to its scalability, it has recently gained popularity and is now winning Kaggle contests for structured data and applied machine learning.

Gradient boosted decision trees (GBM) include an addition called XGBoost that was created specifically to increase speed and efficiency.

4. Results and Discussion:

The following screenshots are depicted the flow and working process of project.

Home page: In our project, we are detecting the fake reviews from the review entered by the user.

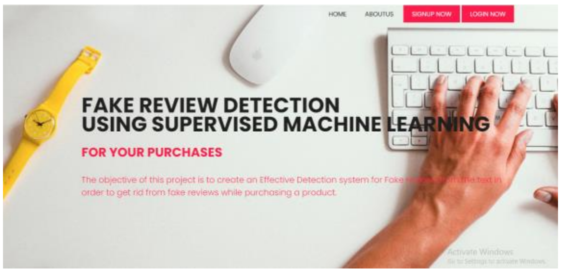


TRUSTED
FAKE REVIEW DETECTOR

FOR YOUR PURCHASES

Detect the fake reviews before making a purchase from e-commerce sites

About page: Here the application describes what main objective of this project is.

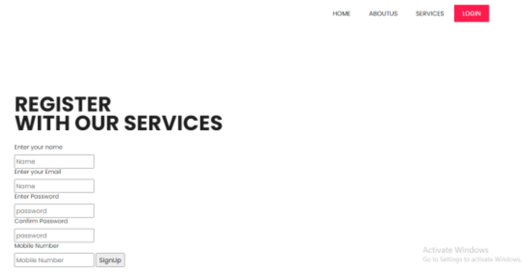


FAKE REVIEW DETECTION
USING SUPERVISED MACHINE LEARNING

FOR YOUR PURCHASES

The objective of this project is to create an Effective Detection system for Fake [illegible] in order to get rid from fake reviews while purchasing a product.

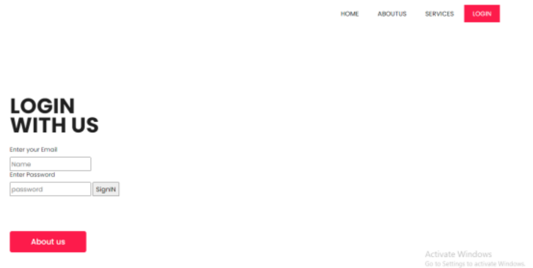
Registration: Registration page in which user need to register to start.



REGISTER
WITH OUR SERVICES

Enter your name
Name
Enter your Email
Name
Enter Password
password
Confirm Password
password
Mobile Number
Mobile Number

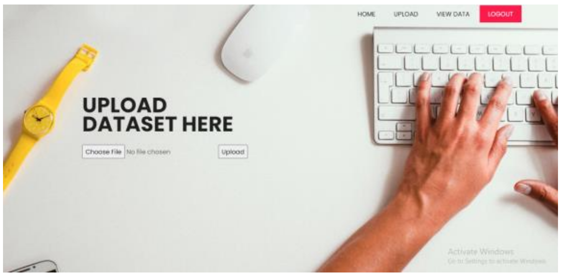
Login page: Here the user need to enter valid credentials in order to enter.



LOGIN
WITH US

Enter your Email
Name
Enter Password
password

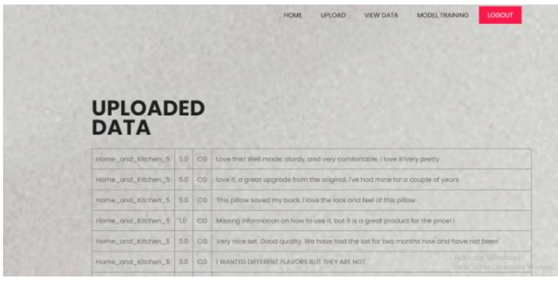
Upload page: Upload Page in order to upload the dataset.



UPLOAD
DATASET HERE

No file chosen

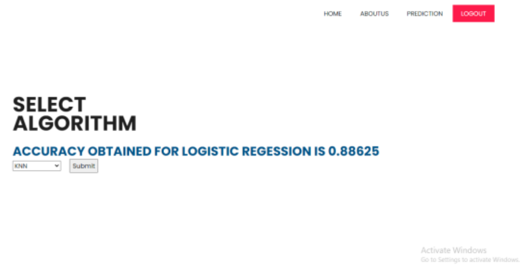
View data: User views the data which he was uploaded to the system.



UPLOADED
DATA

Home_and_Kitchen_5	5.0	CG	Love this! Well made, sturdy, and very comfortable. I love it!Very pretty
Home_and_Kitchen_5	5.0	CG	love it, a great upgrade from the original. I've had mine for a couple of years
Home_and_Kitchen_5	5.0	CG	This pillow saved my back. I love the look and feel of this pillow.
Home_and_Kitchen_5	1.0	CG	Missing information on how to use it, but it is a great product for the price! I
Home_and_Kitchen_5	5.0	CG	Very nice set. Good quality. We have had the set for two months now and have not been
Home_and_Kitchen_5	3.0	CG	I WANTED DIFFERENT FLAVORS BUT THEY ARE NOT.

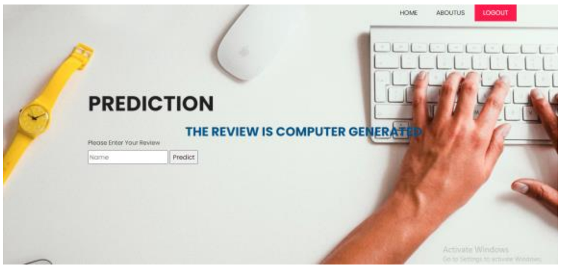
Model training: Here training of your model takes place and display the model's accuracy.



SELECT
ALGORITHM

ACCURACY OBTAINED FOR LOGISTIC REGESSION IS 0.88625

Prediction: Whether the review is machine generated or original, the user must fill out the necessary fields in order to receive a response from the data.



PREDICTION

THE REVIEW IS COMPUTER GENERATED

Please Enter Your Review
Name

V. Conclusion

In this application, we have successfully created a mechanism to identify false reviews. With Python programming and the Django framework, this is produced in a user-friendly environment. In order to establish whether or not the review is fraudulent, the system is likely to collect data from the user.

VI. REFERENCES

[1]. R. Barbado, O. Araque, and C. A. Iglesias, "A framework for fake review detection in online consumer electronics retailers," Information Processing & Management, vol. 56, no. 4, pp. 1234 – 1244, 2019.

[2]. S. Tadelis, "The economics of reputation and feedback systems in e-commerce marketplaces," IEEE Internet Computing, vol. 20, no. 1, pp. 12–19, 2016.

[3]. M. J. H. Mughal, "Data mining: Web data mining techniques, tools and algorithms: An overview," Information Retrieval, vol. 9, no. 6, 2018.

[4]. C. C. Aggarwal, "Opinion mining and sentiment analysis," in Machine Learning for Text. Springer, 2018, pp. 413–434.

[5]. A. Mukherjee, V. Venkataraman, B. Liu, and N. Glance, "What yelp fake review filter might be doing?" in Seventh international AAAI conference on weblogs and social media, 2013.

[6]. N. Jindal and B. Liu, "Review spam detection," in Proceedings of the 16th International Conference on World Wide Web, ser. WWW '07, 2007.

[7]. E. Elmurngi and A. Gherbi, Detecting Fake Reviews through Sentiment Analysis Using Machine Learning Techniques. IARIA/DATA ANALYTICS, 2017.

[8]. V. Singh, R. Piryani, A. Uddin, and P. Waila, "Sentiment analysis of movie reviews and blog posts," in Advance Computing Conference (IACC), 2013, pp. 893–898.

[9]. A. Molla, Y. Biadgie, and K.-A. Sohn, "Detecting Negative Deceptive Opinion from Tweets." in International Conference on Mobile and Wireless Technology. Singapore: Springer, 2017.

[10].S. Shojaee et al., "Detecting deceptive reviews using lexical and syntactic features." 2013.

[11].Y. Ren and D. Ji, "Neural networks for deceptive opinion spam detection: An empirical study," Information Sciences, vol. 385, pp. 213–224, 2017.

[12].H. Li et al., "Spotting fake reviews via collective positive-unlabeled learning." 2014.

Cite this article as :

M. Prathap Reddy, G. Lakshmikanth, "Identification of Fake Reviews Using Supervised Machine Learning", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN : 2456-3307, Volume 8 Issue 6, pp. 116-124, November-December 2022.
Journal URL : https://ijsrcseit.com/CSEIT228610