Fake Product Review Monitoring

A

Project Report

Submitted for the partial fulfilment

of B.Tech. Degree

in

COMPUTER SCIENCE & ENGINEERING

by

Anushka Kanaujia (1805210010)

Astha Sachan(1805210015)

Prateeksha Tiwari(1805210037)

Under the supervision of

Prof. Maheshwari Tripathi

Mr. Mahima Shanker Pandey

Department of Computer Science and Engineering

Institute of Engineering and Technology

Dr. A.P.J Abdul Kalam Technical University, Lucknow, Uttar Pradesh.

May, 2022

Contents

Declaration

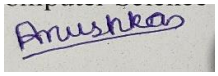
We hereby declare that this submission is our own work and that, to the best of our belief and knowledge, it contains no material previously published or written by another person or material which to a substantial error has been accepted for the award of any degree or diploma of university or other institute of higher learning, except where the acknowledgement has been made in the text. The project has not been submitted by us at any other institute for requirement of any other degree.

Submitted by: - Date:

(1) Name: Anushka Kanaujia

Roll No.: 1805210010

Branch: Computer Science and Engineering

Signature: 

(2) Name: Astha Sachan

Roll No.: 1805210015

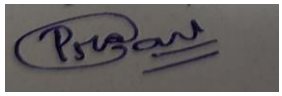
Branch: Computer Science and Engineering

Signature: 

(3) Name: Prateeksha Tiwari

Roll No.: 1805210037

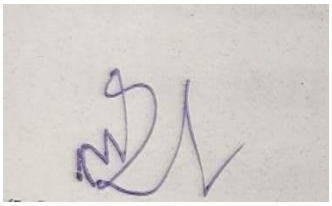
Branch: Computer Science and Engineering

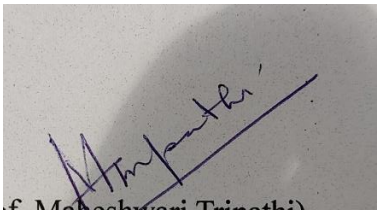
Signature: 

Certificate

This is to certify that the project report entitled “Fake Product Review Monitoring” presented by Anushka Kanaujia, Astha Sachan and Prateeksha Tiwari in the partial fulfilment for the award of Bachelor of Technology in Computer Science and Engineering, is a record of work carried out by them under my supervision and guidance at the Department of Computer Science and Engineering at Institute of Engineering and Technology, Lucknow.

It is also certified that this project has not been submitted at any other Institute for the award of any other degrees to the best of my knowledge.





(Mr. Mahima Shanker Pandey) (Prof. Maheshwari Tripathi)

Department of Computer Science and Engineering

Institute of Engineering and Technology, Lucknow

Acknowledgement

We would like to express our deepest appreciation to our supervisor Prof. Maheshwari Tripathi and co-supervisor Mr. Mahima Shanker Pandey and for giving us this opportunity to engage in a Artificial intelligence based project, we would also like to thank our Head of Department Prof. Divakar S.Yadav , the project committee Prof. Pawan Kumar Tripathi, Prof. Manik Chandra, Prof. Girish Chandra, Ms. Pratibha Pandey and Ms. Deepa Verma as well as for providing us with feedbacks and suggestions to make this project possible. Without their guidance, this project would not have been possible.

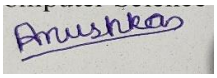
A special acknowledgement goes to our classmates who helped us in completing the project by exchanging interesting ideas and sharing their experience.
We wish to thank our parents as well for their undivided support and interest who inspired us and encouraged us to go our own way, without whom we would be unable to complete our project.

In the end, We want to thank our friends who displayed appreciation for our work and motivated us to continue our work.

(1) Name: Anushka Kanaujia

Roll No.: 1805210010

Branch: Computer Science and Engineering

Signature: 

(2) Name: Astha Sachan

Roll No.: 1805210015

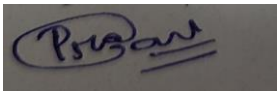
Branch: Computer Science and Engineering

Signature: 

(3) Name: Prateeksha Tiwari

Roll No.: 1805210037

Branch: Computer Science and Engineering

Signature: 

Abstract

In the recent time due to covid-19, online shopping has increased exponentially. In online shopping websites reviews play an integral part in online shopping and booking systems. Showing the right and accurate reviews to user is very important to gain user's trust and help businesses grow.

Our review monitoring system makes 6 checks to check for the negative reviews in large datasets. The six kinds of checks are -
Review which have dual view
Reviews in which same user promoting or demoting a brand
Reviews in which same IP address is promoting or demoting a brand
Reviews posted as flood by same user
Similar reviews posted in same time frame
Meaning less texts in reviews using LSA

The system makes use of :-

Sentimental analysis is a technique that take natural language processing (NLP) into consideration for determining the constructive, deconstructive, or neutral nature of material. In order to assist businesses, textual data is regularly subjected to sentiment analysis.
Latent Semantic Analysis is a type of natural language processing(NLP) that use a qualitative approach to determine the relationship between words and sentences in a document. LSA is tasked with dealing with the types of problems as followed.
The cosine of the angle between two vectors projected in a multi-dimensional space is measured by **cosine similarity**, which is a metric. The closer two vectors are to each other, the smaller the angle between them.

This monitoring system can be used in shopping sites, hotel booking systems, etc to filter out the negative reviews upto a certain accurate extent. Accuracy can be further increased by working on the algorithms used and providing as much training data as possible

Chapter 1
Introduction

In recent times, product reviews on online shopping sites perform a significant role in product sales since people and organizations strive to learn all of the benefits and drawbacks of a product before purchasing it because there are numerous options for the same thing, as there can be different multiple manufacturers who manufacture the same type of product. There could be a variation in the sellers who provide the product, or there could be a difference in the procedure that is followed while making a purchase of the product, so the reviews can be directly linked with the product's sales, and thus it is crucial for online services to filter out fake reviews since their own reputation is at stake. Thus, we need a Fake Review Detection System is needed to discover any suspicious reviews because it's impractical for them to manually check for every review linked with products. So a technology is utilised to try to detect any tendency in the customer reviews.

Many people now days buy and sell products on various e commerce sites and online marketplaces which is which is why demand is growing significantly in number. As a result very informatic feedbacks from customers are also present in these popular sites to help users in examine all the products they are looking forward to buy for their worth.

At one end it’s a very useful and powerful tool in hand but on the bitter side I can sometimes can also lead to wrong decision for users as users can blast these review sections with fake and uttermost opinions which can affect product image in good or bad ways. That’s why it needs to be taken care of as it can be done by seller to rise the popularity of certain product or by opponent or haters to downgrade image of the same product which is of great concerns.

More Commonly, the reviews can be categorized as forgery or authentic review. It’s a pattern when it comes to fake review that reviews are same or more upgraded review for various products. This replication can be separated into four broad categories -

1. Reviews which are posted as flood by same user in which all the reviews are either bad or good.

2. Reviews which are posted in a very large numbers by single person from same IP Address.

3. Reviews in which single user demoting or promoting a particular brand.

4. Reviews in which person with single IP Address demoting or promoting a particular brands image.

The suggested system will save the effort and time by helping the customers and business organizations identify spams from different perspectives and also help the users in purchasing

the right products which will eventually lead to an increase in the user's trust towards the organization. It is very important to deploy a robust and reliable detection algorithm to assure the genuineness of reviews posted on a site. The amount and influence of online reviews is steadily expanding as the Internet is growing in size and importance at a very fast rate.

Reviews can affect people in a very large number of industries in a great number of ways, but the most major industry is e-commerce, where reviews and comments on services and products are often the most easy and comfortable, including many ways for everyone buys to choose if we want to buy a product or not.

Tasks of this project is performed in steps accordingly:

(1) Email will be used to login for better verification purpose.

(2) product features will be mined that have been commented on by customers.

(3) Deciding if a comment is good or bad by getting opinion of sentences in every single review.

(4) If opinions are fake then delete review.

(5) Finally result will be review section with minimum fake reviews

Chapter 2
Literature Review

It is very strenuous to find a spam or factitious reviews. The unauthorised reviewers provide ineligible reviews to the products in order to expand or reduce the sales in less amount of time. There are following styles of reviews present in a shopping website:

1.Dogmatic fake reviews:-

The reviews which contain a large number of views of a person or a user which could be deleterious for the dealing of the product.

>Constructive reviews:-Reviews of this type provide false or ineligible reviews and ratings in order to get increment in the selling of the products

>Deconstructive reviews:-Reviews of this type are posted in order to create a decrement in the purchase of the product.

2. Reviews on brands/Websites only:-

These comments aren't about the product, but about the company or website that sells it. These types of reviews are unhelpful and are only designed to increase the product's quality sales.

3. Unrelatable reviews:-

These types of reviews aren't qualified to be termed reviews because they aren't about the product. They may be a personal experience or some gibberish text

Previously, some mechanisms or projects were designed to address the issue of bogus reviews.

1.Detection of spams using ratings behaviour[1] :-

When it comes to spotting spams or fake reviews, it's vital to look at the ratings' behaviour. There's a potential that spammers would spam the product for irrational events, even if the product itself is fine. In this case, the overall ratings or average ratings will come to the rescue.

2.Spotting a group of fake reviewers[2]:-

A large number of reviews that are fraudulent and are made to the same product are considered spam or fraudulent since they might be placed to degrade or belittle the product sales.

3.Online reviews manipulation[3]:-

Because spammers may easily change ratings in numeric or textual formats, it is critical to assess reviews in graphic or picture format, which will determine the good and bad features of the evaluations, as well as their legitimacy. Hence we will not be able to find it the review is authentic or not.

4. A Review on fake product review detection and removal techniques [4]:-

Because reviews are so important in today's internet world, where online commerce is rising by leaps and bounds, a tool to detect bogus reviews is required. As a result, many approaches for removing fraudulent reviews are presented in this study.

Between 1942 and 1962, the period was characterised by a confluence of technological discoveries and advancements (of which the Second World War was an accelerator or drive) and the desire to understand how machines and humans interact. The goal, according to Norbert Wiener, a pioneer in cybernetics, was to combine mathematical theory, electronics, and automation as "a comprehensive theory of control and communication, both in animals and machines." As early as 1943, Warren McCulloch and Walter Pitts created the first mathematical and computer model of the biological neuron (formal neuron).

The fresh rise in the discipline around 2010 can be attributed to two things.

1. First and foremost, access to massive amounts of data. In order to use algorithms for picture classification and cat recognition, for example, it used to be necessary to sample oneself. As early as 1943, a model of the biological neuron (formal neuron) was created.
2. Then, to boost the speed of the computation of learning algorithms, the extremely high efficiency of computer graphics card processors were invented. Because the procedure is iterative and exhausting, processing the complete sample might take weeks before 2010.

Opinion mining did not begin with the introduction of the internet and Web 2.0, as one might imagine, but rather far earlier. As a result, pinpointing the exact moment when opinion mining began is extremely difficult.

The majority of the publications and research from 1981 to 1996, the initial phases, are text explanation. The majority of studies in this era highlighted the importance of interpreting and categorising text into opinion using a simple algorithm. This step entails significantly relying on human interaction and psychological qualities to analyse the text.

The second phase, from 1998 to 2004, shows some improvement, with researchers focusing not just on classification but also on the development of annotation tools in text and documents. Endnotes and footnotes are two examples of successful computerised text annotation systems for enhanced attribute recognition in text processing.

The third phase, from 2004 to 2006, focuses on interpretation and extraction approaches. Papers from this era show that researchers are discovering and developing ways and algorithms to retrieve documents from the internet.

Using classification algorithms, researchers attempted to discriminate between facts and opinions in the materials. The genuine spirit of opinion mining began in phase 4 in 2007, when a wave of new applications in industries such as politics and business, among others, were produced.

A comparative study of the past researches and the project is present in the following table:

Detection of spams using ratings behavior [1](2010)	- Checked rating behaviour of the product. - Unnecessary bad or good reviews were eliminated
Spotting a group of fake reviewers [2](2012)	- Spotted group of spammers that are posting same reviews multiple time using same IP address
Online reviews manipulation[3](2015)	- To find out the genuine review graphics analysis was done, but as each user does not post picture with review this was not very effective.
A Review on fake product review detection and removal techniques[4](2021)	- Mentions the six ways to treat the fake reviews - Discuss which one of six is the best
Fake Product Review Monitoring(our Project) (2022)	- Our project incorporate all the six steps in a well organized manner so that the reviews at the end will mostly be genuine.

Chapter 3
Motivation

In today's era of online shopping and e-commerce, online reviews play a quite important role in decision-making. Customers, for example, check product or store reviews before determining what to buy, where to buy it, and whether or not to buy it. Because there are monetary incentives provided to produce false/ fraudulent reviews, there has been a major surge in difficult opinion spam on online review websites. In essence, an untruthful review is a phoney, fraudulent, or opinion spam review. Positive reviews and ratings on a specific product can attract more customers and increase sales; bad evaluations might reduce demand and sales. In recent years, fake review detection has received a lot of attention.

However, most review sites still do not publicly filter bogus reviews.

Disadvantages of existing systems.

> User is not able to find out whether the review is genuine or fake and can make a wrong purchase reading fake reviews which leads to the waste of time and money.

> If the optimization team posts the reviews from a different IP address, the system will be unable to detect the fake reviews.

> Brands can use their resources to wrongly increase the rating of their particular products.

> Same user can write multiple reviews from different accounts.

Advantages of proposed system.

> Users get maximum genuine reviews about the product.

> Users can post their own review about the product.

> User can save their time and money by shopping through websites which have genuine reviews of customers.

> Fraudulent cases can be detected and taken care of.

Websites such as Amazon, Flipkart, Myntra, etc. sell numerous products. It is very usual to sometimes see that a product has multiple reviews which appear to be same. This in turn creates a bad user experience while shopping products and destroys the usefulness of the reviews. The existing system can be integrated with such websites to create a better user experience

Chapter 4

Methodology

In recent years, online reviews have become increasingly important in purchase decisions. Client assessments may give a plethora of information about your product or service. These may not be accurate and spammers or frauds will be able to fake it and manufacture phoney reviews if there is a fictitious, or drop in, the quality of the products or services. Buyers will be misinformed as just a consequence of spammers' actions, and they'll still make poor decisions all of the time. As a result, detecting spam opinions is a significant issue. Spam is defined as the employment of excessive and illegal means, such as the fabrication of a huge number of false opinions, whether favourable or negative, in order to generate positive or negative reviews. A technique for detecting false product reviews is based mostly on mining industry assessments, making product purchases more reliable for our consumers.

The general concept is to take product information from a dataset and extract reviews from there, then pass these reviews through a false review detection model to filter out fake and actual reviews, and then remove the phoney reviews from the page[5]. We can utilise artificial intelligence in a fake review detection system/model, which will have a variety of methos to detect content according to user requirements, and data is the key to all of this.

Artificial intelligence, abbreviated as AI, is software that enables computers to act and think like humans. According to some developers, artificial intelligence can perform as well as or better than humans. We define performance as the precision, speed, and capacity of human computing. Artificial intelligence (AI) is a branch of computer science that investigates and develops computer systems capable of doing tasks that need human intelligence. Artificial intelligence includes speech recognition, decision-making, and visual perception. The ability to translate between languages is another useful function.

Customer reviews on a product will be used as data to which procedures will be applied. Mining method can be used to further break down the phoney review monitoring system and sift the false reviews from the genuine ones. This is attributable to the fact that consumer feedback can provide a wealth of information concerning your service or product. However, in order to improve the image or reduce the quality of the products or services spammers can fake reviews. Customers will be fooled as a consequence of spammers' activity, and they will all make poor decisions most of the time. Therefore, the detection of a spam opinions, it is a serious problem. A methodology for recognizing phony customer reviews is based primarily on mining industry assessments, making product transactions more trustworthy for our customers.

The overall idea is to take product information from the dataset and extract review from there then when have reviews with us in the form of text, will pass these review from a fake review detection model and filter out fake and real reviews and will remove the fake review .For a fake review detection system/model we can use artificial intelligence which will have a variety of methods to detect content according to user requirements and key to all this is data.

Artificial intelligence, sometimes known as AI, refers to software technologies that enable computers to behave and think like humans. According to some developers, artificial intelligence can perform as well as or better than humans. "Performance" relates to the accuracy, speed, and capacity of human computation. Artificial intelligence (AI) is a branch of computer science that studies and develops computer systems that can perform activities that would normally need human intellect. Speech recognition, decision-making, visual perception, and other human cognitive characteristics may be possessed by artificial intelligence. Another aspect is language translation.

Here reviews given by customers on a product will act as data on which will apply methods. To further break down the fake review monitoring system for filtering the fake reviews from a genuine one, can carry out a mining method. mining. Through comprehensive data analytics, mining has improved corporate decision-making. the technique of extracting patterns and other useful data from big data collections. There are many mining methods like:

1. Text Mining (has been widely used in knowledge-based organizations. The technique of reviewing enormous amounts of documents in order to find new information is known as text mining.)
2. Opinion mining and sentiment analysis (a technique of analysis that employs computer linguistics and natural language processing to automatically identify and extract sentiments and reviews from text (positive, negative, neutral, etc.))
3. Natural language processing or NLP, is an acronym that stands in the broadest sense of the word, and is defined as the automated processing of natural language as a language test.

Fig 1. Architecture of project

Here for mining fake reviews out of all can use sentiment analysis where The most commonly used method of NLP is sentiment analysis. Sentiment analysis is most useful in cases such as this, as the client, advisory, research, reviews, and comments on the social networks, where people can express their opinions and feedback. The easiest way to do it is to analyze sentiment in order to make get sentiment like a positive/negative/neutral). In more complex cases, as the result of a numerical score, which can be divided into as many categories as you want. In the case of our sample text, and the customer can clearly articulate the feelings in different parts of the text. Because of this, the output is not very useful. Instead, we can find the meaning of a sentence and to separate the positive and negative parts of the review. The Sentiment is from the review it can also help you choose the best of the positive and negative portions of the review.

4.1 **Machine Learning Techniques Used**

Cosine Similarity

Cosine similarity is a metric that computes the cosine of the angle formed by two vectors projected in three dimensions.

Two vectors are similar to each other if the angle between the two vectors is smaller.

If the angle between two vectors is 90 degrees, the cosine similarity will have a value of 0; this means that the two vectors are perpendicular to each other which means they are not correlated.

The angle between vectors A and B decreases as the cosine similarity measurement approaches 1. In this scenario, it is possible to say that A and B are more similar.

Cosine similarity can be described mathematically as the division between the dot product of vectors and the product of the euclidean norms or magnitude of each vector.

$\text{Cos}(\theta)$ lies in the range $[-1,1]$:

-1 - indicates highly opposite vectors i.e. no similarity

0 - indicates independent vectors

1 - indicates a high similarity between the vectors

Some applications of cosine similarity are:

In processes of data mining, information retrieval, and text matching

Is used in a recommendation engines to recommend similar entities such as books, clothes, etc

The cosine-similarity based locality-sensitive hashing technique increases the speed for matching DNA sequence data.

Latent Semantic Analysis

Latent Semantic Analysis (LSA) is a natural language processing method that use statistical techniques to find associations between words in a document. It deals with issues such as:

For example, mobile, phone, cell phone, and telephone are all comparable, but if we ask "The cell phone has been ringing," only documents with the word "cell phone" are returned, but documents with the words "mobile, phone, and telephone are not returned.

LSA makes the following two assumptions:

1. Words used in the same context are comparable to one another.
2. The ambiguity of the terms used obscures the data's latent semantic structure.

Language is more than just the words on the page in front of you. When you read a text, your mind creates images and ideas in your head. Themes develop after reading a large number of works, even if they are never mentioned directly. Our ability to comprehend and absorb language defies a mathematical description (for the moment). It is one of the most widely used Natural Language Processing (NLP) strategies for quantitatively determining text topics.

Chapter 5

Proposed Model

The fake product review monitoring system collects reviews from various users and detects fraudulent reviews using sentiment analysis/opinion mining and content-similarity approaches, assisting the user in purchasing the proper products based on genuine customer reviews. The method will also assist in spotting customers who write repeated reviews with the intent of harming the brand or company's reputation. This model will helps us to detect the fake reviews and treat them

The following steps will be taken to complete the task:

1. The customer's e-mail address will be confirmed during login;
2. Gathering feedback on product features that have been mentioned by customers.
3. Determine if each comment is positive or negative by identifying opinion sentences in each review.
4. If fraudulent reviews are discovered while offering opinions, a warning will be sent and the review will be designated as a phoney review.
5. Summarizing the findings by review section, with the least amount of bogus reviews possible.

The system consists of the:

Frontend – The frontend part of the system will consist of a website where users will be able to create their google account and will be authenticated through Google Auth. On successful login of a particular user, the website will show the various products available. On the purchase of the particular product, the user will be able to submit his/her review about the respective product.

Backend – The backend part of the system will be responsible for storing the reviews of the particular product along with the details of the user who made the purchase of the respective product.

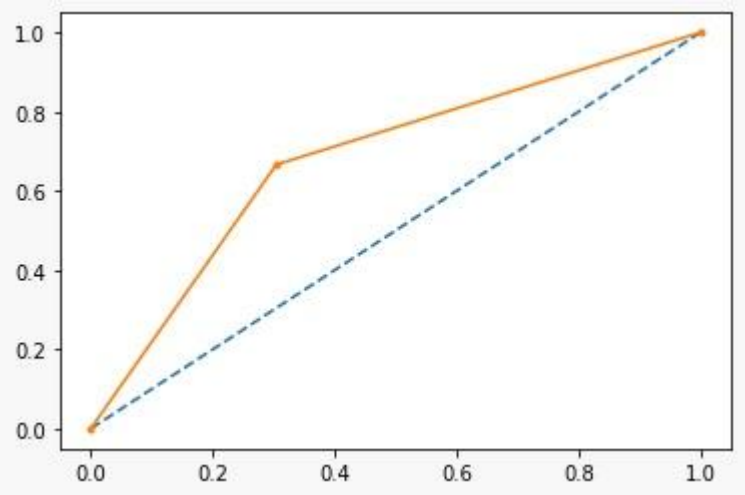
Review Monitoring System- The review monitoring system will be responsible for detecting the type of review whether positive or negative through sentiment analysis (opinion mining), content similarity to detect the multiple reviews given by the same user to detect the frauds.

Sentimental Analysis- For sentimental analysis use of the certain algorithms will be done. Also focus on which classifier has the most accuracy.

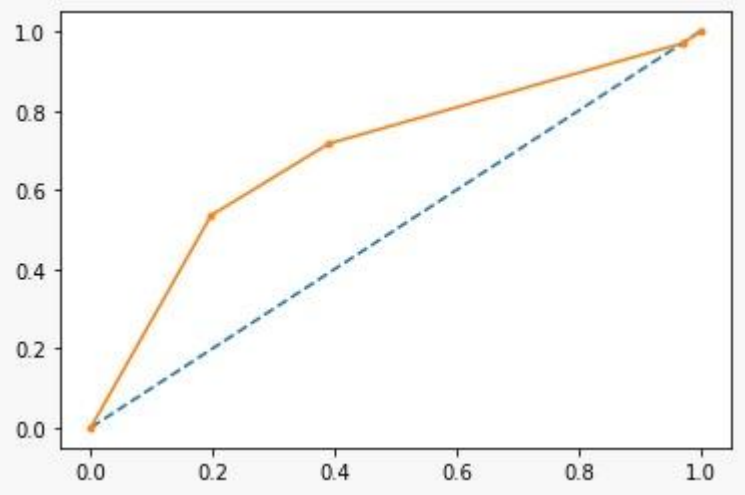
To identify similar material from our dataset for redundant data, we used a content similarity technique. Content similarity is a statistic which can be used to evaluate how similar data objects are, regardless of their size. Cosine Similarity is a Python function that measures the similarity of two sentences. Data items in a dataset are handled as a vector in cosine similarity, which has the following benefits: Even if the two similar data objects are separated by the Euclidean distance due to their size, the angle between them could be smaller. The greater the similarity, the smaller the angle. The cosine similarity captures the orientation (angle) of the data items when plotted on a multi-dimensional space, not the magnitude.

Model Accuracy at various Stages for data set

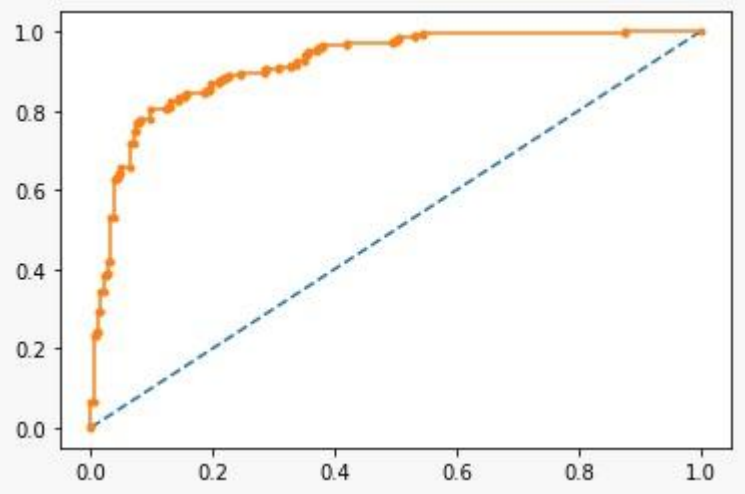
x-axis- probability of positive outcome
y-axis- Accuracy



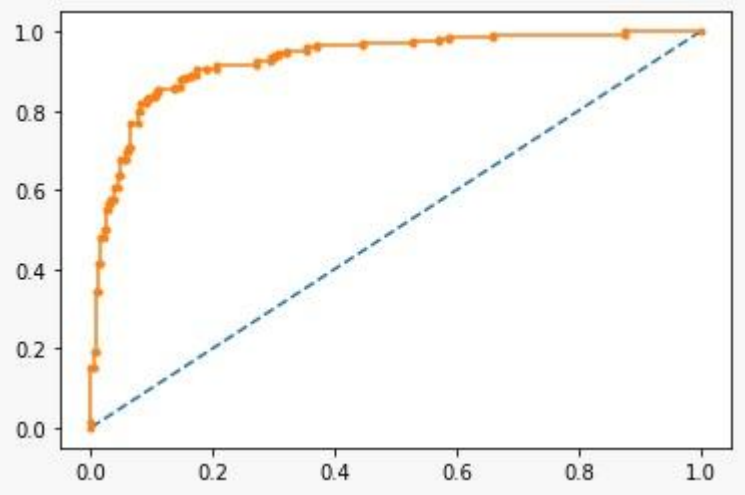
Initial Phase (Graph 1)



Mid Phase (Graph 2)



Pre-Final Phase (Graph 3)



Final Phase (Graph 4)

6 Steps for filtering out fake reviews

1. Review which have dual review

Dual reviews refer to those reviews which have different sentiments in their heading and body which means that the heading is positive regarding the product and the body is negative or vice versa. These reviews are filtered out by matching the sentiment of the heading and the body of the review. If the sentiment is not found equal, such reviews are marked as fake.

2. Reviews in which same user is promoting or demoting a brand.

In this step, reviews on a certain brand products are collected through the user id of the person and then checked for multiple reviews

3. Reviews in which same IP address is promoting or demoting a brand

There are cases when the optimisation team try to bombard the review section from different IP addresses to defame other brands. In this step we group reviews based on the same IP address and mark them as fake for a particular product.

4. Reviews posted as flood by same user

This step is quite simple. Here we handle cases where a particular user posts multiple reviews on the same product which is not at all required. These steps are done intentionally to defame a product. Such reviews are marked as fake by the system.

5. Similar reviews posted in same time frame

Sometimes people use bots which generate multiple reviews and flood the review section. These reviews can vary in the text. So here we have used a time frame method where reviews posted in a short time frame by a particular user on same or different products are marked as fake.

6. Meaningless text in reviews using LSA

The last step is to handle meaningless reviews. Some users tend to write their own meaningless story which has nothing in relation with the product. To filter out such reviews we have used latent semantic analysis

Chapter 6
Plan of Work

MONTHS	WORK TARGET
January 2022	• Worked on the idea and synopsis of the project • Assessed various methodologies through the published research papers
February 2022	• Design the front end of the website • Authenticate users using Google OAuth • Add about us and contact page
March 2022	• List products on the website • Include the backend part and database to store user reviews • Detect multiple reviews submitted through the same IP addresses
April 2022	• Inclusion of sentimental analysis, latent semantic analysis, and content similarity. • Final testing of system • Further enhancements on accuracy

System Requirements

Hardware Requirement -

The following describes the hardware needed in order to execute and develop the fake product review monitoring:

● Computer Desktop or Laptop

The computer desktop or a laptop will be utilized to run the software in Order to run application in system. A notebook which is a small, lightweight and inexpensive laptop computer is proposed to increase mobility.

System will be using

OS: Windows 7 or later one

Processor : Core2Duo

Main Memory : 4-16 GB RAM

Hard Disk : 520GB

Display : 14.5" Monitor

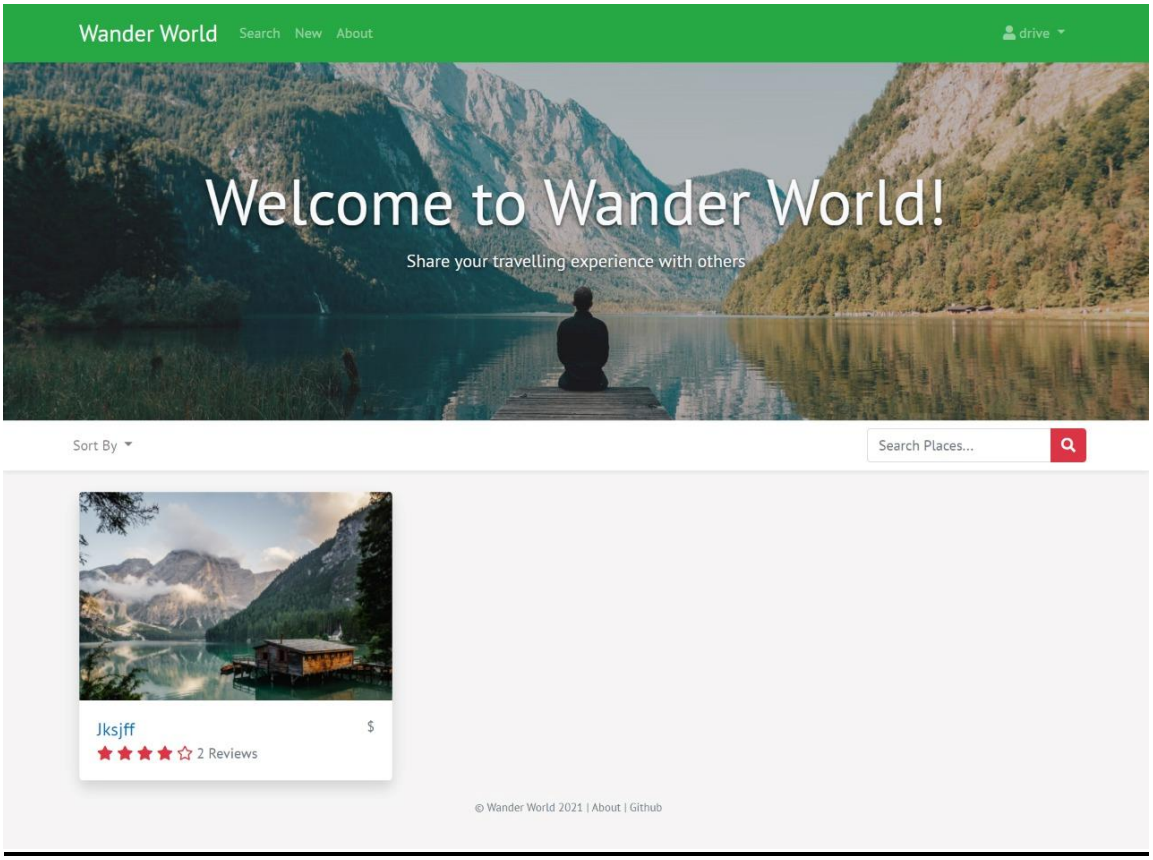
Software Requirement -

The following describes the software needed in-order to develop the Virtual Mouse application:

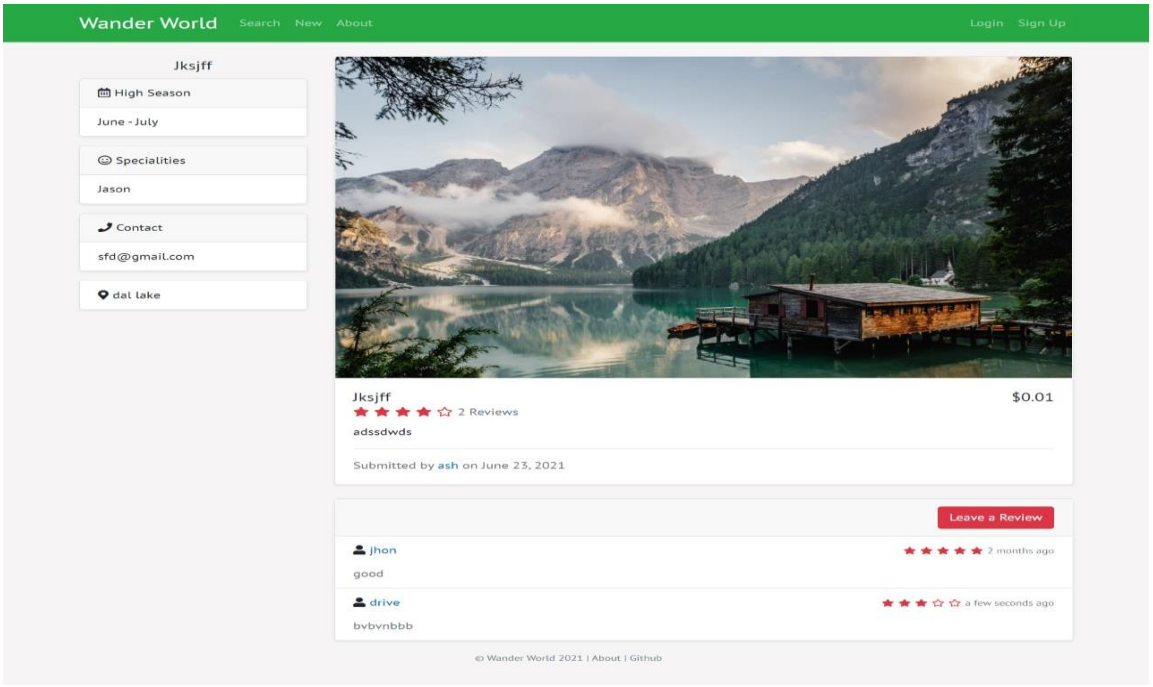
● Python Language:
The coding technique on developing the Virtual Mouse application will be Python with the aid of the integrated development environment (IDE) that are used for developing computer programs, known as the Microsoft Visual Studio. A Python library provides many operators, covering basic arithmetic, bit manipulation, indirection, comparisons, logical operations and others.

● Jupyter Notebook:
The Jupyter Notebook is the original web application for creating and sharing computational documents. It offers a simple, streamlined, document-centric experience.

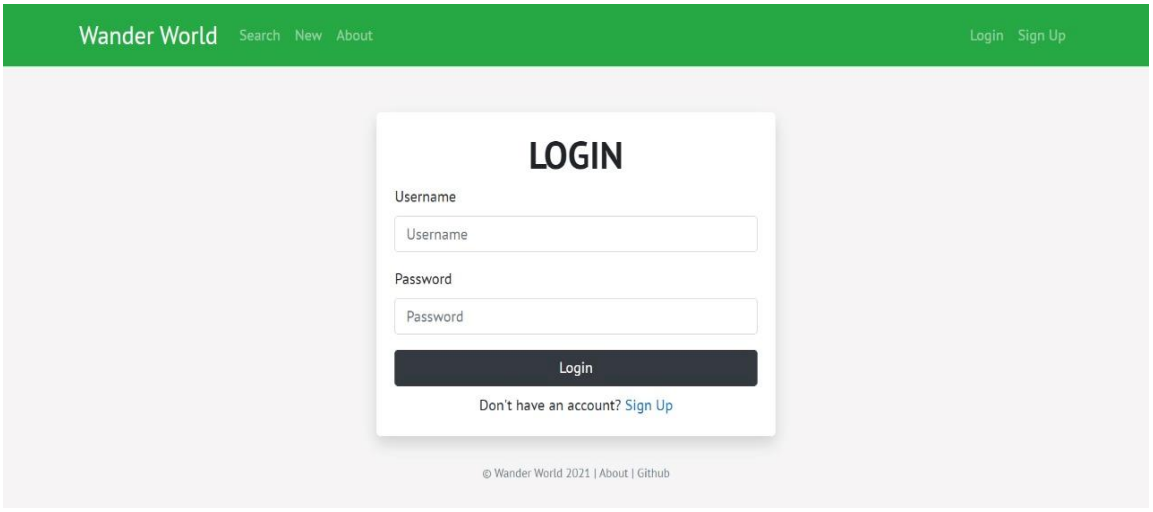
SNAPSHOTS



Home Page

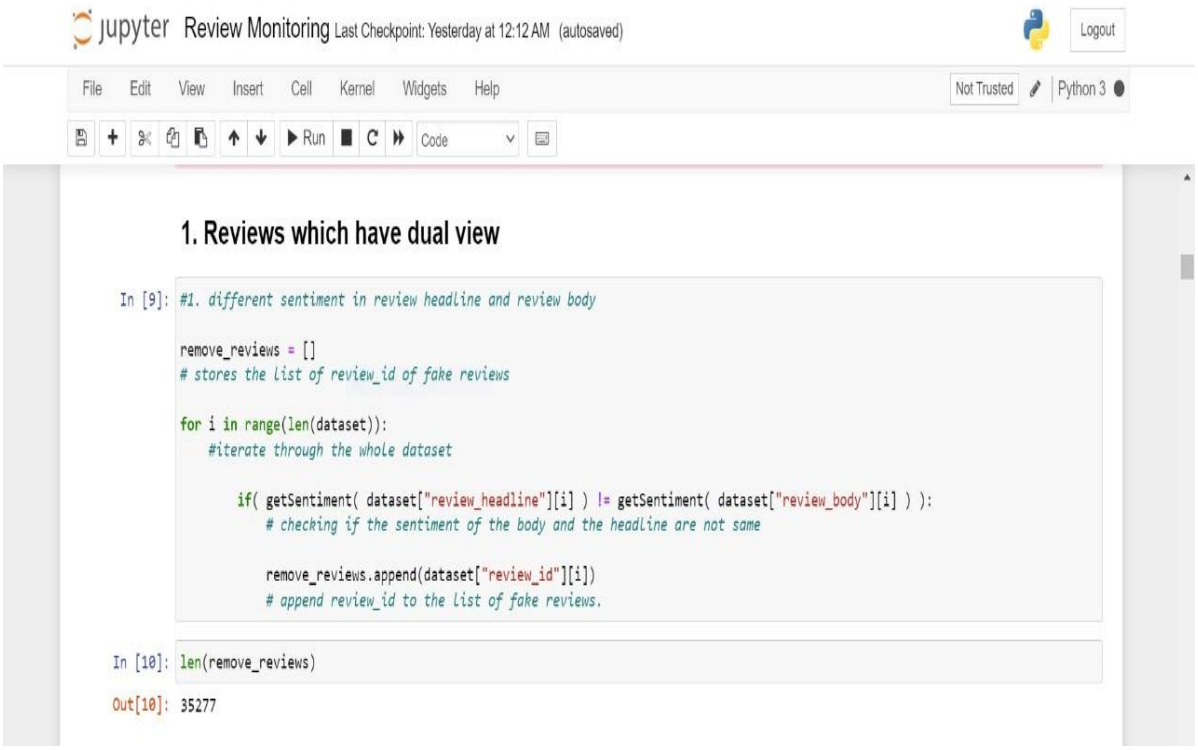


Reviews Page



Login Page

Code Outputs



2. Reviews in which same user promoting or demoting a particular brand

In [11]:
```
#2. Users which are posting either all positive or negative reviews on different products of same brand

customers = dataset.groupby("customer_id")
# groups dataset by customers

customer_list = dataset["customer_id"].unique()
#list of unique customers

size = len(customer_list.tolist())
#size of total unique customers

for i in range(size):
    # iterate through all the customers

    brand_df = customers.get_group(customer_list[i])
    # Dataframe for each customers

    brands = brand_df.groupby("product_parent")
    # groups reviews of each customers by brand

    brands_list = brand_df["product_parent"].unique()
    # unique list of brands for each customers reviews

    no_of_brands = len(brands_list.tolist())
    # no. of brands for which reviews had been written by the customer

    for j in range(no_of_brands):
        # iterate through all the brands

        product_df = brands.get_group(brands_list[j])
        # Dataframe of products for a brand for which a customer had written reviews

        no_of_products = len(product_df["product_id"])
        # no of products

        if no_of_products<=2:
            # it will filter the products which are less than 2 for a brand
            continue

        indices = product_df.index.values.tolist()
        # index of the dataframe of the products of each brand for each customers

        sentiment = getSentiment(product_df["review_body"][indices[0]])
        # sentiment of the review of the first product

        isSameSentiment = True

        #discarding those cases in which we have only less than 5 reviews on same brand
        if(no_of_products<5):
            continue

        for k in range(1,no_of_products):
            # iterate through all the products

            text = str(product_df["review_body"][indices[k]])
            # review of each product

            if getSentiment(text)!=sentiment :
                # if sentiment is different than discard it
                isSameSentiment = False
                break;

        if(isSameSentiment):
            # if sentiments of all the products of same brand by a customer is same,
            #append customer_id to blocked users list

            remove_reviews.append(customer_list[i])
            break
```

In [12]:
```
len(remove_reviews)
```

Out[12]:
```
35278
```

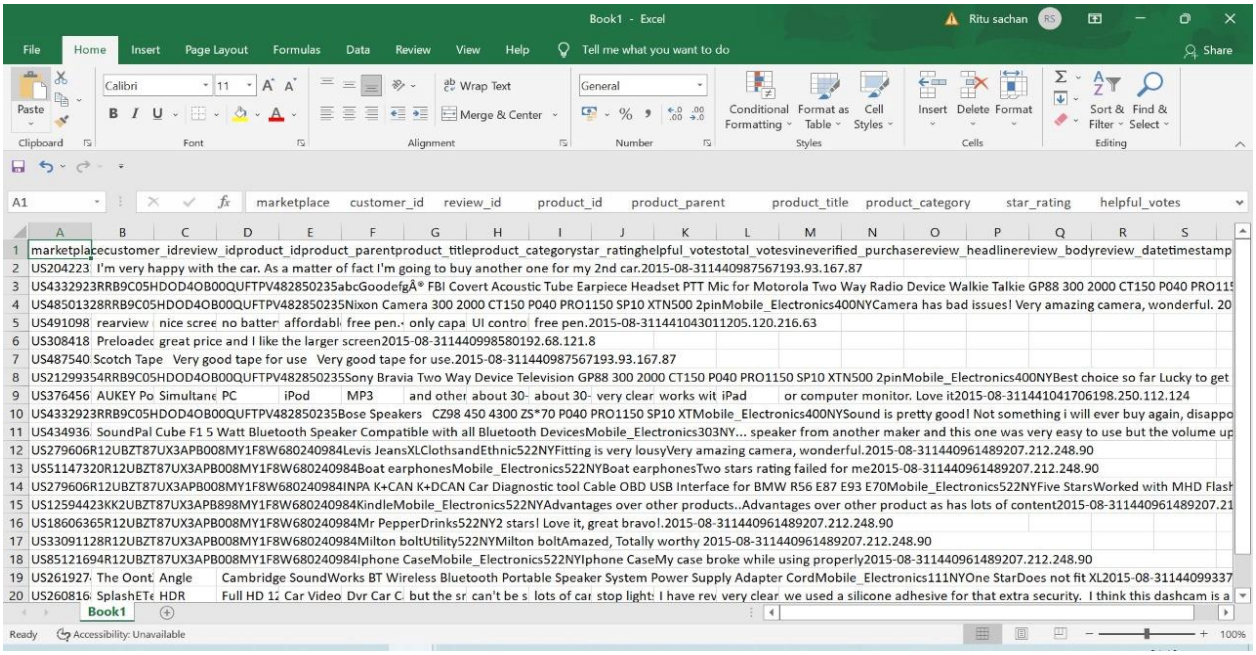
3. Reviews in which person from same IP Address promoting or demoting a particular brand

In [13]:
```
#3.Reviews from same IP either all positive or negative reviews on different products of same brand


ip = dataset.groupby("IP Address")
#grouping the dataset by ip address

ip_list = dataset["IP Address"].unique()
#stores the list pf unique ip addresses

remove_ip = []
#stores the list of ip address from where reviews have been written.

size = len(ip_list.tolist())
#stores the size of the total unique ip addresses

for i in range(size):
    # iterate through all the ip addresses

    brand_df = ip.get_group(ip_list[i])
    # Dataframe of brands for which reviews have been written from the same ip address

    brands = brand_df.groupby("product_parent")
    # grouping the products of the same brands for each ip addresses

    brands_list = brand_df["product_parent"].unique()
    #list of unique brands for each ip addresses

    no_of_brands = len(brands_list.tolist())
    # total no. of brands

    for j in range(no_of_brands):
        # iterate through all the brands

        product_df = brands.get_group(brands_list[j])
        # Dataframe of the products of each brand of each products

        no_of_products = len(product_df["product_id"])
        # no of products of each brand for each ip addresses

        if no_of_products<=2:
            # filter the reviews of the brandswith less than 3 reviews
            break

        indices = product_df.index.tolist()
        # indices of dataframe of products of each brand for each customers

        sentiment = getSentiment(product_df["review_body"][ indices[0] ])
        # sentiment of review of first product of each brand

        isSameSentiment = True

        for k in range(1,no_of_products):
            # iterate through all the reviews

            text = str(product_df["review_body"][indices[k]])
            # reviews of each product

            if getSentiment(text)!=sentiment :
                # if sentiment of 2 products of same brand are not same
                # then check the next brand
                isSameSentiment = False
                break;

        if(isSameSentiment):
            # if all the sentiments are same , append ip to blocked list
            remove_ip.append(ip_list[i])
```

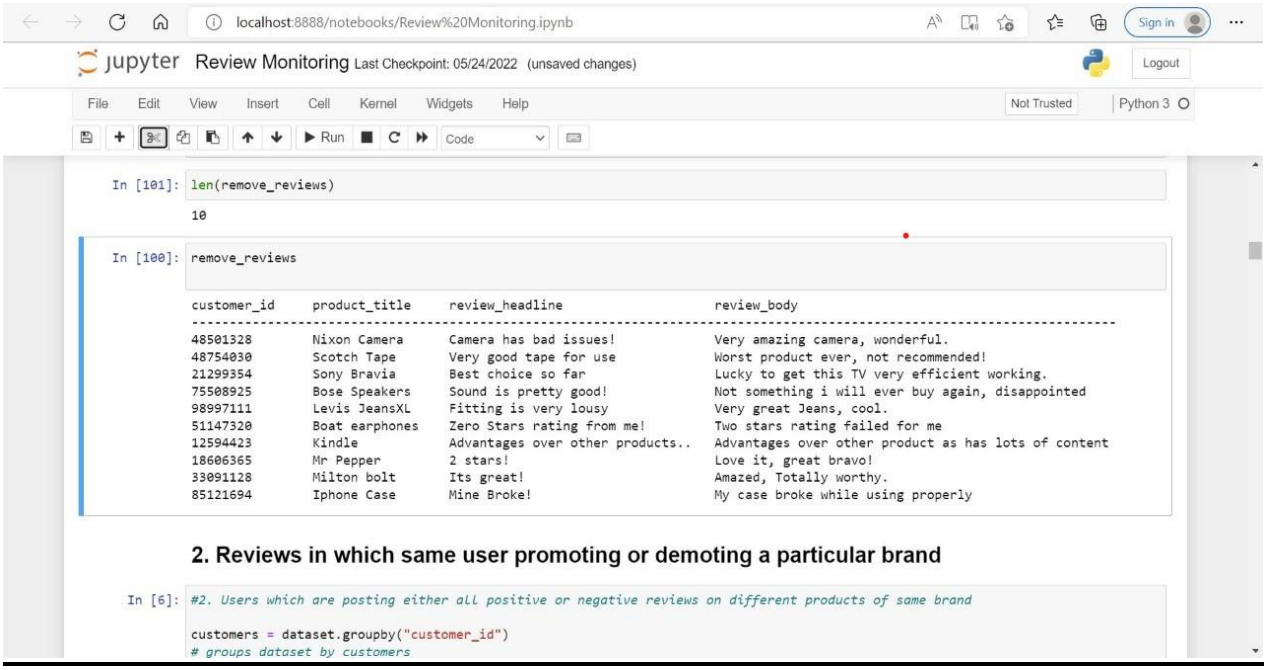
In [14]:
```
remove_ip
```

Out[14]:
```
['193.93.167.87']
```

Dataset



Fake reviews output



Conclusion

Detecting opinion spam from large amounts of unstructured data has become a significant research challenge as a result of this study. Although various algorithms have been utilised in opinion review analysis and have yielded positive results, no particular algorithm can address all of the obstacles and difficulties that today's systems faces. Our programme will assist the user in purchasing the appropriate product without falling into the trap of any scams. For genuine ratings, people can acquire a report on Fake Product Review Monitoring & Removal. Our application will analyse the data and then post real product reviews. Also, the consumer can be certain upto a certain extent that the products are available with genuine reviews.

Our main objective is to develop a system that can detect spam and duplicated reviews and filter them out, providing users with reliable information regarding the product. Our project's goal is to improve customer satisfaction while also making online buying more secure. By using opinion mining techniques and constructing a word dictionary, the project will be able to detect false reviews.

It is feasible to enhance the algorithm used to calculate review sentiment scores.. It is possible to update our sentiment word dictionary. Possibility of adding more terms to our lexicon and updating the weights assigned to those words in order to obtain a more accurate review score

References

[1] Duan, Detecting product review spammers,2010-Ee-Peng Lim(School of Information Systems Singapore Management University) https://www.academia.edu/2724864/Detecting_product_review_spammers_using_ rating_behaviors

[2] Spotting Fake Reviewers using Product Review Graph,2015-Zhou Wang(School of Information Science and Engineering, Shenyang Ligong University, Shenyang , China https://www.academia.edu/23397297/Spotting_Fake_Reviewers_using_Product_Review_Graph

[3] Manipulation of online reviews: An analysis of ratings, readability, and sentiments, 2012, by Noi Sian Koh https://www.academia.edu/18916989/Manipulation_of_online_reviews_An_analysis_of_ratings_readability_and_sentiments

[4] A Review on fake product review detection and removal techniques,2021-Rutuja B. Ardak, Prof. Girish S.Thakre (Sipna C.O.E.T, Amravati, India)https://ijcrt.org/papers/IJCRT2108104.pdf

[5] https://towardsdatascience.com/sentiment-analysis-concept-analysis-andapplications-6c94d6f58c17

[6] https://www.aviso.com/blog/sentiment-analysis-using-nlp/

[7] https://kavita-ganesan.com/what-is-text- 5 similarity/

[8] http://www.cs.wisc.edu/niagara/data/.

[9] https://www.kernix.com/article/similarity-measure-of-textual-documents/

[10] https://towardsdatascience.com/an-easy-tutorial-about-sentiment-analysis-withdeep-learning-and-keras-2bf52b9cba91.

[11] https://cfm.ehu.es/ricardo/docs/python/Learning_Python.pdf

[12] Effective JavaScript: 68 specific ways to harness the power of JavaScript by David Herman 5 December 2012.

[13] http://ajbasweb.com/old/ajbas/2013/April/359-370.pdf