INTERNET TECHNOLOGY AND APPLICATIONS (CO368)

Fake Product Review Monitoring and Product Evaluation using Opinion Mining

FINAL REPORT

Submitted to:

Saumya Hegde

Sushmitha G K

Date: 18-04-2018

Submitted by:

Roll no	Name	Email	Phone
15CO148	Shivananda D	shivananda199@gmail.com	8453543399
15CO154	Yeshwanth R	yeshwanthr98@gmail.com	7760087265
15CO152	Suhag	suhaga15@gmail.com	9880371334

CONTENTS

INTRODUCTION

Today, the world has been taken over by the internet. The technological progress that takes place everyday in this world is tremendous. In this fast-moving technological world, the internet is accessed by almost everyone who owns a smartphone or a desktop computer or a laptop. The internet has made our lives easier by helping us immensely in many ways. Most of the people can complete their work by sitting in one place and accessing the internet. For example, students carry out their projects by referring to online resources and study through online courses, and one of the most common things happening on the internet is the online shopping. People prefer to buy products online rather than going out and buying at the actual shops. Online shopping has grown tremendously. We see millions of orders being delivered to customers' door steps by companies providing online shopping facility. Before buying any product, people look at the ratings/reviews of that product to know whether the product is good or not. Seller selling products on the web often ask or take reviews from customers about the products that they have purchased. As e-commerce is growing and becoming popular day-by-day, the number of reviews received from customers about the product grows rapidly. So, people come across various reviews in the website, but whether those reviews are genuine is not identified by the user buying the product online. For a popular product, the reviews can go up to thousands. This creates difficulty for the potential customer to read them and to decide whether to buy or not buy the product. In some review websites, some good reviews are added by the product company people itself to make product famous. These people give good reviews for many products manufactured by their own firm. So, customers buying products online look at those reviews, think that the product is good and buy them. To find out and remove those fake reviews, this Fake Product Review Monitoring system is introduced. The basic idea is to track the IP address of the user submitting the review and if the same kind of review is coming from identical IP address, then those reviews are removed. The system also rates a product based on the reviews submitted by the customers. The system incorporates machine learning and opinion mining technology to rate a product.

LITERATURE SURVEY

1. **Fake Product Review Monitoring and Removal for Genuine Ratings PHP** by *Manleen Kaur Kohli, Shaheen Jamil Khan, Tanvi Mirashi, Suraj Gupta*: The systems already present today can permit users to submit their reviews and ensure that all reviews are submitted by a legitimate user. But, the contents of the reviews are not analysed by the administrators for identifying fake reviews. According to various studies, 2-6% reviews present online are fake reviews, and these reviews misguide customers. So, it is important to identify and remove those fake reviews. Various methods have been introduced to identify the fake reviews. One method is by looking at the timestamps and contents of the reviews. Multiple reviews with same timestamp can have similar phrases rearranged in the review content, which imply the same positive/negative feedback regarding a product. These can be identified and marked as fake review. In our system, the administrator can look at the content of the reviews and mark the review as fake if it is felt a fake review. But this method is time consuming as it involves analysing the content of multiple reviews to identify whether the review is fake or not. So, our system aims at faster fake review identification and fake review removal.

2. **Sentiment Analysis and Opinion Mining: A Survey** *by G Vinodhini and R M Chandrasekaran:* This has been a very useful survey that we have used as a reference. The existing online shopping systems allow customers to rate a product themselves. Based on number of reviews and number of positive ratings, product recommendation is done. But here also, a manufacturing firm can simply give a five-star rating to their products and advertise their products as good products. Our system aims to overcome this by allowing the administrator to evaluate a product based on user reviews using opinion mining. This is an extra measure of ensuring that fake ratings are not promoted, and customers are not misguided. One of the common methods to rate a product is by doing the sentiment analysis of the product reviews. In this method, the review is analysed, and each word is assigned a weight based on whether the word is indicating a positive sentiment or negative sentiment. For this, a database of positive sentiment and negative sentiment words is maintained. Also, a database containing stopwords such as him, I, me, etc which do not convey any sentimental meaning. These words are removed from the reviews during sentiment analysis. Then, the sentiment weights of each word in the review is added and based on the score, the review is classified as positive or negative. But, this method just recommends a product just based on positive or negative classification of reviews. Our system aims to rate a product on a scale of 5 based on customer reviews. To do this, machine learning and opinion mining techniques are used by our system.

IMPLEMENTATION

To implement our system of Fake Product Review Monitoring and Product Evaluation using Opinion Mining, an Online Shopping site is developed which is hosted and maintained on a local server. Anyone with the IP address of the local server and project directory path, can access the website and enjoy shopping online. The shopping site contains products of five categories (men, women, gadgets, sports and books). The main purpose of the project is to identify and remove fake reviews and implement product evaluation using opinion mining. So, the functionalities and features of the online shopping site are limited but user friendly. The basic steps involved in the implementation of the system are as follows:

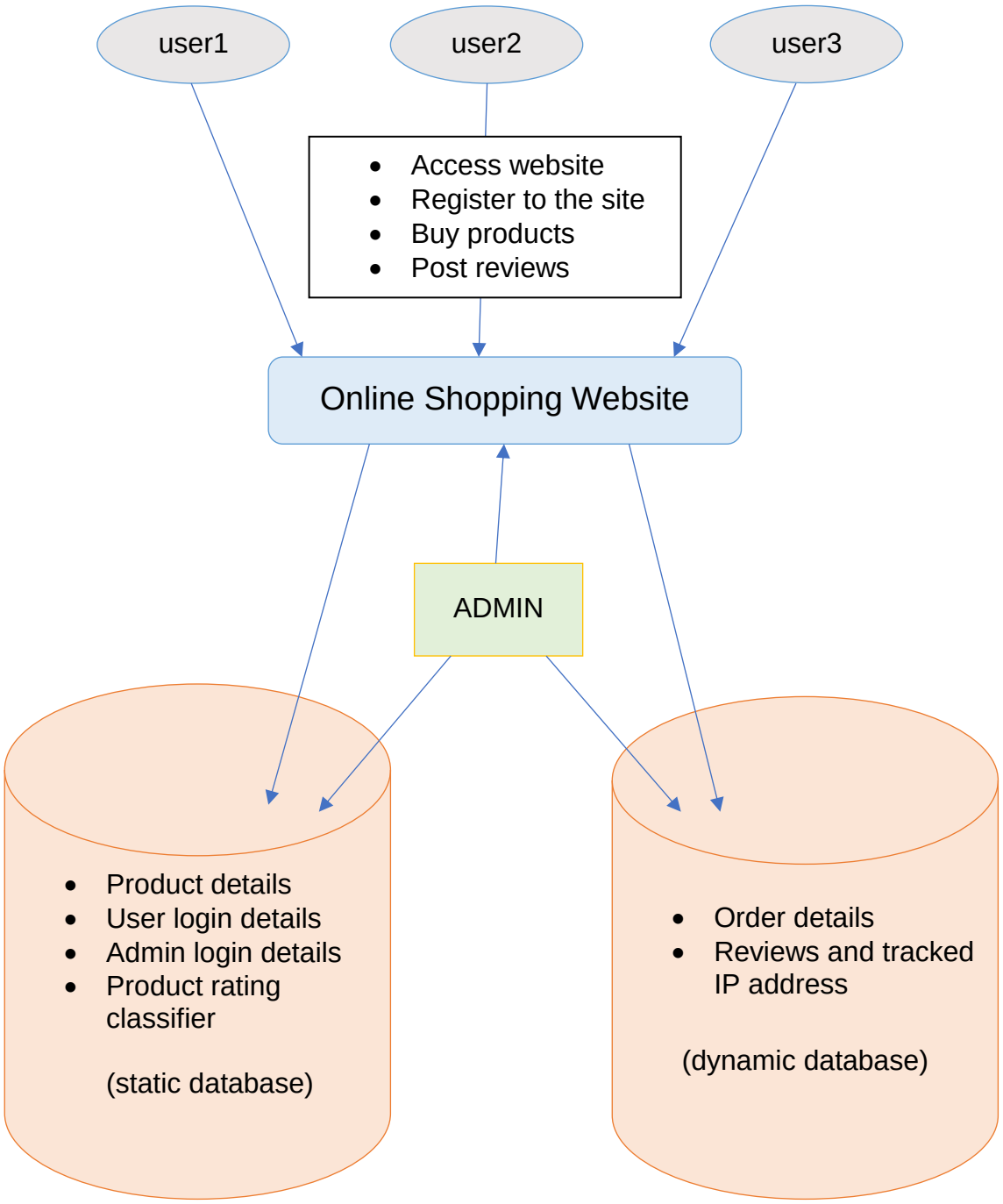
- The Online Shopping website is built using web-scripting languages
- Admin will add the products to the system for the users to shop online
- User wanting to do the online shopping must first register using their email
- Registered users can login, access the website, view products and do the shopping (only registered users can buy the products)
- Users can also submit reviews only on products which they have bought
- The system will track the IP address of users submitting the reviews
- Admin can again add more products to the system or delete a product from the shopping site
- Admin can also view the reviews submitted and remove them if found fake (if multiple reviews are being submitted from same IP address, admin can identify them as fake review and remove them)
- Admin will also rate a product on a scale of 5 based on the reviews submitted by the users using machine learning and opinion mining technology

So, basically there are two primary users of the system. One is the admin maintaining the online shopping site and the database, and the other one being the customers who buy products. Therefore, the system consists of two modules:

Admin Module: where admin will login, add/delete product, identify and remove fake products, rate a product based on customer reviews

User Module: where the online customer will register, buy products and write reviews on the bought products

BLOCK DIAGRAM



PROPOSED SYSTEM

The system is composed of a collection of websites that are accessed by external users and some set of websites which are exclusively available only to system administrators. The websites accessed by the users contains various products that are available for purchase, i.e., an online shopping site. This is a common system among any company offering online shopping facilities. The websites are built using web designing languages like HTML, CSS, JavaScript and Bootstrap. The websites are completely user friendly. Firstly, users are asked to register to the site through their email if they wish to purchase any product available on the site. Only registered users can buy the products.

Any user who has purchased the product can review it. However, the product manufacturers can take advantage of this fact and can post multiple good reviews of their product to promote it and make it popular. But, the product would be very bad. This results fake reviews. To tackle this situation, our system tracks the IP address of the user submitting the review of the product. Then the administrator will go through the database containing the reviews. The database will contain the reviews, user details i.e., the user who submitted the review along with the tracked IP address. The administrator will search for the instances where multiple reviews implying the same meaning have been submitted by the same IP address. These reviews are identified as fake reviews and the administrator will remove those reviews.

The system also rates a product based on the user reviews of the products. This is done using the machine learning and opinion mining techniques. The classification will consist of five classes, representing one for each of 5-star rating i.e., class1 for 1-star rating, class2 for 2-star rating, class3 for 3-star rating, class4 for 4-star rating and class5 for 5-star rating. Firstly, a classifier is built using the ‘sklearn’ machine learning library. The LinearSVC (a linear support vector machine) classifier imported from sklearn library is used as a classifier. The classifier is trained using a huge database containing the reviews and its ratings. The database used to train the classifier was downloaded online which was available freely which contained thousands of amazon product reviews and its ratings. Then the trained classifier was used to rate the product on a scale of 5 using the user reviews. Initially, to train the classifier and to predict the rating of a review, the review must be represented numerically and should be given as input to the classifier. This is accomplished using the TfidfVectorizer imported from sklearn feature extraction library. Also, pre-processing of the review text is done by removing the stopwords (which do not convey any sentimental meaning). This task of pre-processing is accomplished using nltk library which contains English stopwords. The classifier cannot be trained based on single words in review text, because consider the case when the user review is “not good”. This must be given a 2 or 3-star rating since it is a negative review. But if we train based on single words, this will be given a 4 or 5-star rating after recognizing the word “good” in the review. This case is handled by specifying minimum and maximum range for n-gram attribute in TfidfVectorizer. So, if we specify maximum as 2 for n-gram in TfidfVectorizer, the case where a positive word is negated such as “not good” is handled. This is how a product is rated based on the reviews submitted by the users.

TOOLS USED FOR IMPLEMENTATION

- XAMPP tool – provides Apache local server and integrated MySQL database
- Web programming languages HTML, CSS, JavaScript, Bootstrap – to build websites and GUI required
- PHP – for server-side scripting and for tracking IP address
- Python code – for running the product rating system using machine learning classifier

HARDWARE REQUIREMENTS

- Processor: i3
- Hard disk space: 50GB and higher
- RAM: 1GB and higher

SOFTWARE REQUIREMENTS

- Operating system: Windows 7 or higher
- Server: Apache local server (XAMPP tool)
- Database: MySQL (XAMPP tool)

WORK DISTRIBUTION

Websites/GUI (Front-end)	Shivananda D Yeshwanth R Suhag
Admin Module • PHP server-side scripting • Add/delete products • MySQL Database Handling (Back-end) • IP address tracking and fake product review identification and removal • Product rating based on user reviews (text pre-processing, text vectorization, building classifier and predict rating)	Shivananda D
User Module • Add product details to database • User registration/login • Product details webpages	Yeshwanth R Suhag