SENTIMENTAL ANALYSIS OF TWITTER USING MACHINE LEARNING

By

P.GOWTHAMRAJ 2016103024

V.DHANAPRADEEP 2016103019

*Abstract*:

Social media have received more attention nowadays. Public and private opinion about a wide variety of subjects are expressed and spread continually via numerous social media. Twitter is one of the social media that is gaining popularity. Twitter offers organizations a fast and effective way to analyze customers’ perspectives toward the critical to success in the market place. Developing a program for sentiment analysis is an approach to be used to computationally measure customers’ perceptions. This paper reports on the design of a sentiment analysis, extracting a vast amount of tweets. Prototyping is used in this development. Results classify customers’ perspective via tweets into positive and negative, which is represented in a pie chart and html page. However, the program has planned to develop on a web application system, but due to limitation of Django which can be worked on a Linux server or LAMP, for further this approach need to be done.

*Keywords-component; Twitter, sentiment, opinion mining, social media, natural language processing*

I. INTRODUCTION

According to [1], millions of people are using social network sites to express their emotions, opinion and disclose about their daily lives. However, people write anything such as social activities or any comment on products. Through the online communities provide an interactive forum where consumers inform and influence others.Moreover, social media provides an opportunity for business that giving a platform to connect with their customers such as social media to advertise or speak directly to customers for connecting with customer’s perspective of products and services. In contrast, consumers have all the power when it comes to what consumers want to see and how consumers respond. With this, the company’s success & failure is publicly shared and end up with word of mouth. However, the social network can change the behavior and decision making of consumers, for example, [2] mentions that 87% of internet users are influenced in their purchase and decision by customer’s review. So that, if organization can catch up faster on what their customer’s think, it would be more beneficial to organize to react on time and come up with a good strategy to compete their competitors

*A. Problem Statement*

Despite the availability of software to extract data regarding a person’s sentiment on a specific product or service,organizations and other data workers still face issues regarding the data extraction.

Sentiment Analysis of Web Based Applications Focus on Single Tweet Only.

With the rapid growth of the World Wide Web, people are using social media such as Twitter which generates big volumes of opinion texts in the form of tweets which is available for the sentiment analysis [3]. This translates to a huge volume of information from a human viewpoint which make it difficult to extract a sentences, read them, analyze tweet by tweet, summarize them and organize them into an understandable format in a timely manner [3].

* Difficulty of Sentiment Analysis with inappropriate English. Informal language refers to the use of colloquialisms and slang in communication, employing the conventions of spoken language [4] such as ‘would not’ and ‘wouldn’t’. Not all systems are able to detect sentiment from use of informal language and this could hanker the analysis and decision-making process.

Emoticons, are a pictorial representation of human facial expressions [5], which in the absence of body language and prosody serve to draw a receiver's attention to the tenor or temper of a sender's nominal verbal communication, improving and changing its interpretation [6]. For example indicates a happy state of mind. Systems currently in place do not have sufficient data to allow them to draw feelings out of the emoticons. As humans often turn to emoticons to properly express what they cannot put into words [6]. Not being able to analyze this puts the organization at a loss. Short-form is widely used even with short message service (SMS). The usage of short-form will be used more frequently on Twitter so as to help to minimize the characters used. This is because Twitter has put a limit on its characters t o 1 4 0 [ 7 ] . F o r e xam p l e , ‘Tba’ refers to be announced**.**

*B. Objective*

The objectives of the study are first, to study the sentiment analysis in microblogging which in view to analyze feedback from a customer of an organization’s product; and second, is to develop a program for customers’ review on a product which allows an organization or individual to sentiment and analyzes a vast amount of tweets into a useful format.

II. METHODOLOGY

This project has been divided into 2 phases. First, literature study is conducted, followed by system development. Literature study involves conducting studies on various sentiment analysis techniques and method that currently in used. In phase 2, application requirements and functionalities are defined prior to its development. Also, architecture and interface design of the program and how it will interact are also identified. In developing the Twitter Sentiment Analysis application, several tools are utilized, such as Python Shell 2.7.2 and Notepad.

LITERATURE REVIEW :

*A. Opining Mining*

Opinion mining refers to the broad area of naturallanguage processing, text mining, computational linguistics, which involves the computational study of sentiments, opinions and emotions expressed in text [8]. Although, view or attitude based on emotion instead of reason is often colloquially referred to as a sentiment [8]. Hence, lending to an equivalent for opinion mining or sentiment stated that opinion mining has many application domains including accounting, law, research, entertainment, education, technology, politics, and marketing. In earlier days many social media have given web users avenue for opening up to express and share their thoughts and opinions [10].

*B.Twitter*

Twitter is a popular real time microblogging service that allows users to share short information known as tweets which are limited to 140 characters [2,3], [11]. Users write tweets to express their opinion about various topics relating to their daily lives. Twitter is an ideal platform for the extraction of general public opinion on specific issues [9,10]. A collection of tweets is used as the primary corpus for sentiment analysis, which refers to the use of opinion mining or natural language processing [1].

Twitter, with 500 million users and million messages per day, has quickly became a valuable asset for organizations to invigilate their reputation and brands by extracting and analyzing the sentiment of the tweets by the public about their products, services market and even about competitors [12]. [2] highlighted that, from the social media generated opinions with the mammoth growth of the world wide web, super volumes of opinion texts in the form of tweets, reviews, blogs or any discussion groups and forums are available for analysis, thus making the world wide web the fastest, most comprising and easily accessible medium for sentiment analysis.

*E.Twitter Sentiment Analysis*

The sentiment can be found in the comments or tweet to provide useful indicators for many different purposes [20]. Also, [12] and [36] stated that a sentiment can be categorized into two groups, which is negative and positive words. Sentiment analysis is a natural language processing techniques to quantify an expressed opinion or sentiment within a selection of tweets [8].

Proposed work:

Machine learning methods often rely on supervised classification approaches where sentiment detection is framed as a binary which are positive and negative [24]. This approach requires labeled data to train classifiers [21]. This approach, it becomes apparent that aspects of the local context of a word need to be taken into account such as negative (e.g. Not beautiful) and intensification (e.g. Very beautiful) [19]. However, [20] showed a basic paradigm for create a feature vector is:

* Apply a part of speech tagger to each tweet post
* Collect all the adjective for entire tweet posts
* Make a popular word set composed of the top N adjectives
* Navigate all of the tweets in the experimental set to create the following:
* Number of positive words ,Number of negative words,Presence, absence or frequency of each word

Installation:

* Tweepy: [tweepy](http://docs.tweepy.org/en/v3.5.0/) is the python client for the official [Twitter API](https://dev.twitter.com/rest/public).  
  Install it using following pip command:

pip install tweepy

* TextBlob: [textblob](http://textblob.readthedocs.io/en/dev/) is the python library for processing textual data.  
  Install it using following pip command:

pip install textblob

Also, we need to install some NLTK corpora using following command:

python -m textblob.download\_corpora

(Corpora is nothing but a large and structured set of texts.)

pip install matplotlib : helps to create the graph in python

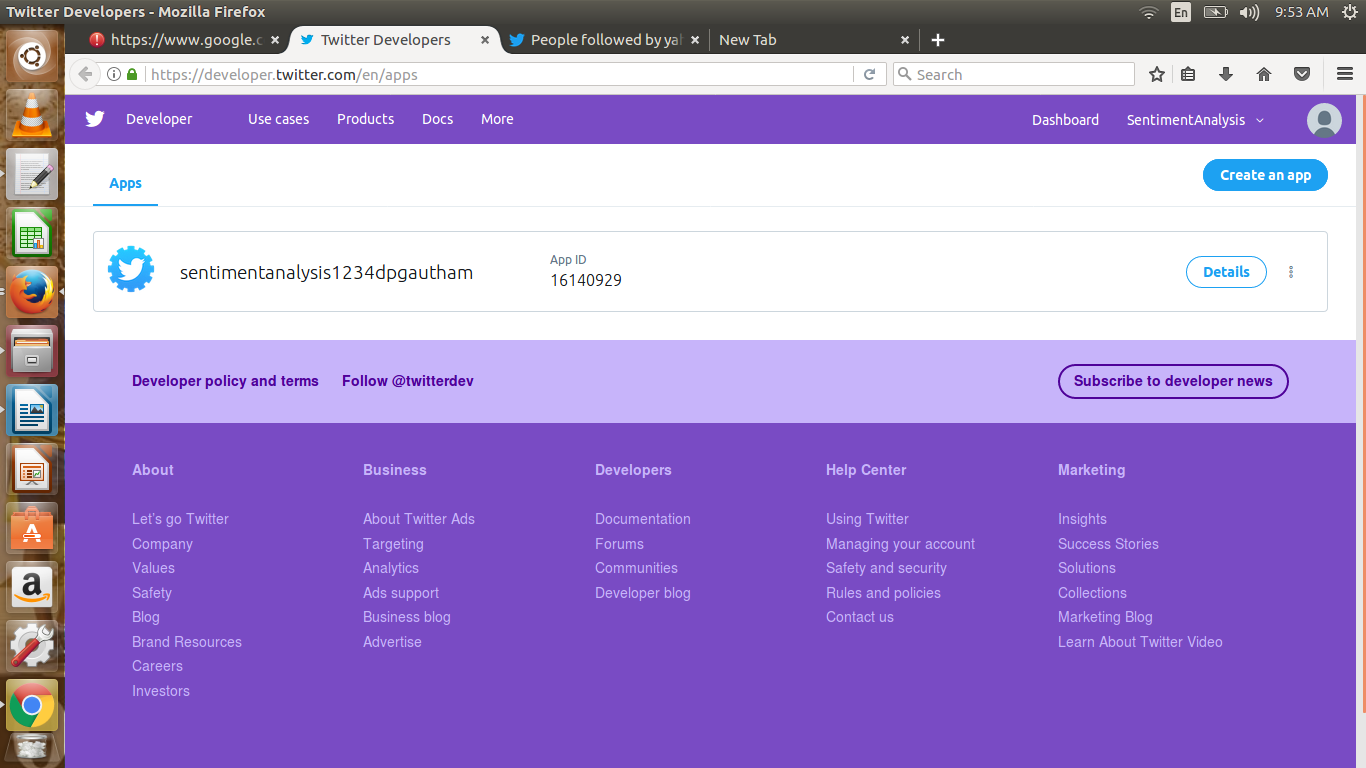
Modules :

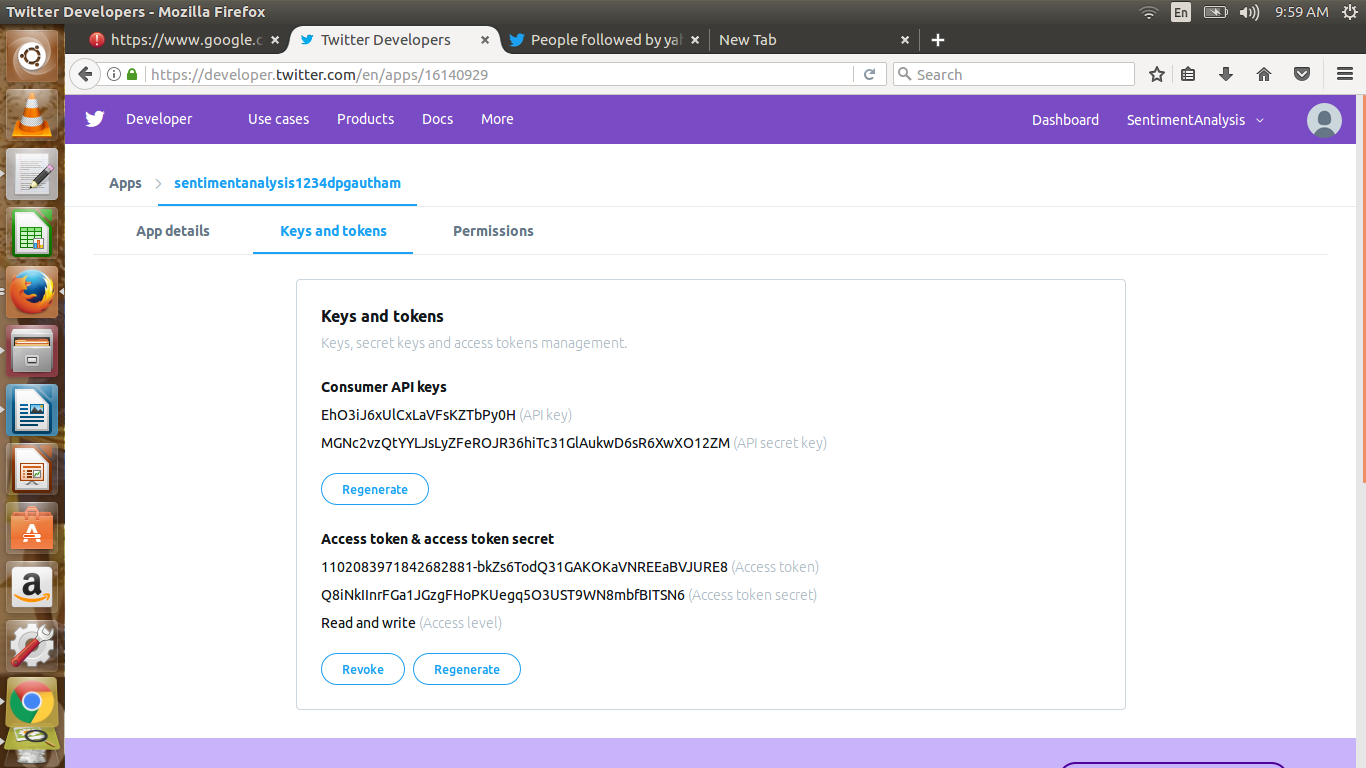
1)App Creation: Inorder to livestream the twitter we have to create the app in the twitter developer website .

Fill the application in the site and finally twitter review the process and let you create the app of your own

link:<https://developer.twitter.com/en/apps>

your created app looks like below:



 fig : our twitter app for sentimental analysis

2)Authenication:

The twitter app generates the consumer and access token keys .copy and paste it the python code inorder to perform anlaysis.

Fig: our consumer and access token keys generated in the twitter app

The above keys are highly confidential since it associated with the twitter accout.i.e,if the analysis performed against the twitter rules,the twitter account is going to be locked.So the keys should be secret to the individual.

Steps:

In order to fetch tweets through Twitter API, one needs to register an App through their twitter account. Follow these steps for the same:

* Open this  [link](https://apps.twitter.com/) and click the button: ‘Create New App’
* Fill the application details. You can leave the callback url field empty.
* Once the app is created, you will be redirected to the app page.
* Open the ‘Keys and Access Tokens’ tab.
* Copy ‘Consumer Key’, ‘Consumer Secret’, ‘Access token’ and ‘Access Token Secret’.

3) fetch tweets:

Get the keyword and count of tweets to be fetched from the user

save the tweets in the file format .csv

code :

# input for term to be searched and how many tweets to search

searchTerm = input("Enter Keyword/Tag to search about: ")

NoOfTerms = int(input("Enter how many tweets to search: "))

# searching for tweets

self.tweets = tweepy.Cursor(api.search, q=searchTerm, lang = "en").items(NoOfTerms)

# Open/create a file to append data to

csvFile = open('result.csv', 'a')

# Use csv writer

csvWriter = csv.writer(csvFile)

4) Polarity:

Fix all the polarity to the value zero.

Based on the below condition the analysis is performed.

if (analysis.sentiment.polarity == 0): # adding reaction of how people are reacting to find average later

neutral += 1

elif (analysis.sentiment.polarity > 0 and analysis.sentiment.polarity <= 0.3):

wpositive += 1

elif (analysis.sentiment.polarity > 0.3 and analysis.sentiment.polarity <= 0.6):

positive += 1

elif (analysis.sentiment.polarity > 0.6 and analysis.sentiment.polarity <= 1):

spositive += 1

elif (analysis.sentiment.polarity > -0.3 and analysis.sentiment.polarity <= 0):

wnegative += 1

elif (analysis.sentiment.polarity > -0.6 and analysis.sentiment.polarity <= -0.3):

negative += 1

elif (analysis.sentiment.polarity > -1 and analysis.sentiment.polarity <= -0.6):

snegative += 1

5) Report generation:

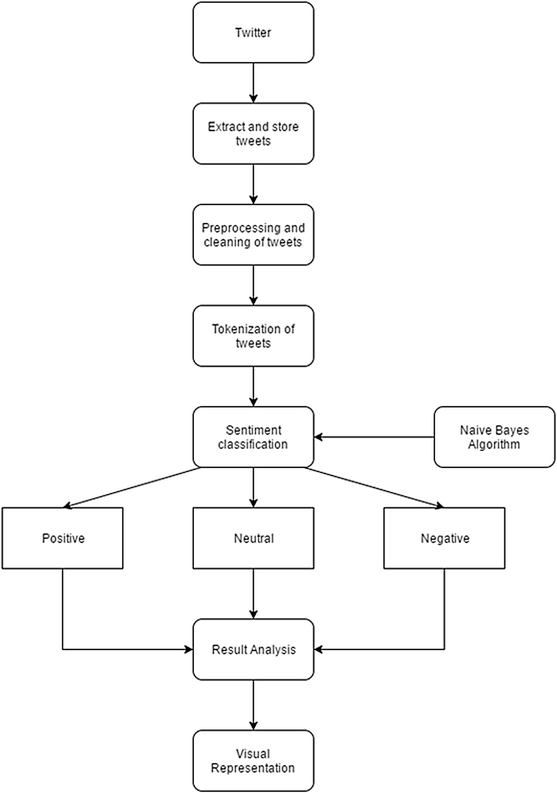
positive = self.percentage(positive, NoOfTerms)

Like the above method it is possible to calculate the percentage of people reacting on the particular topic

After that generate the detailed report on the peoples reaction towards the tweet

6)Graph generation:

plt.show() method in python helps to geneate the piechart graph Different color shows the different opinion of people towards the tweets



Output:

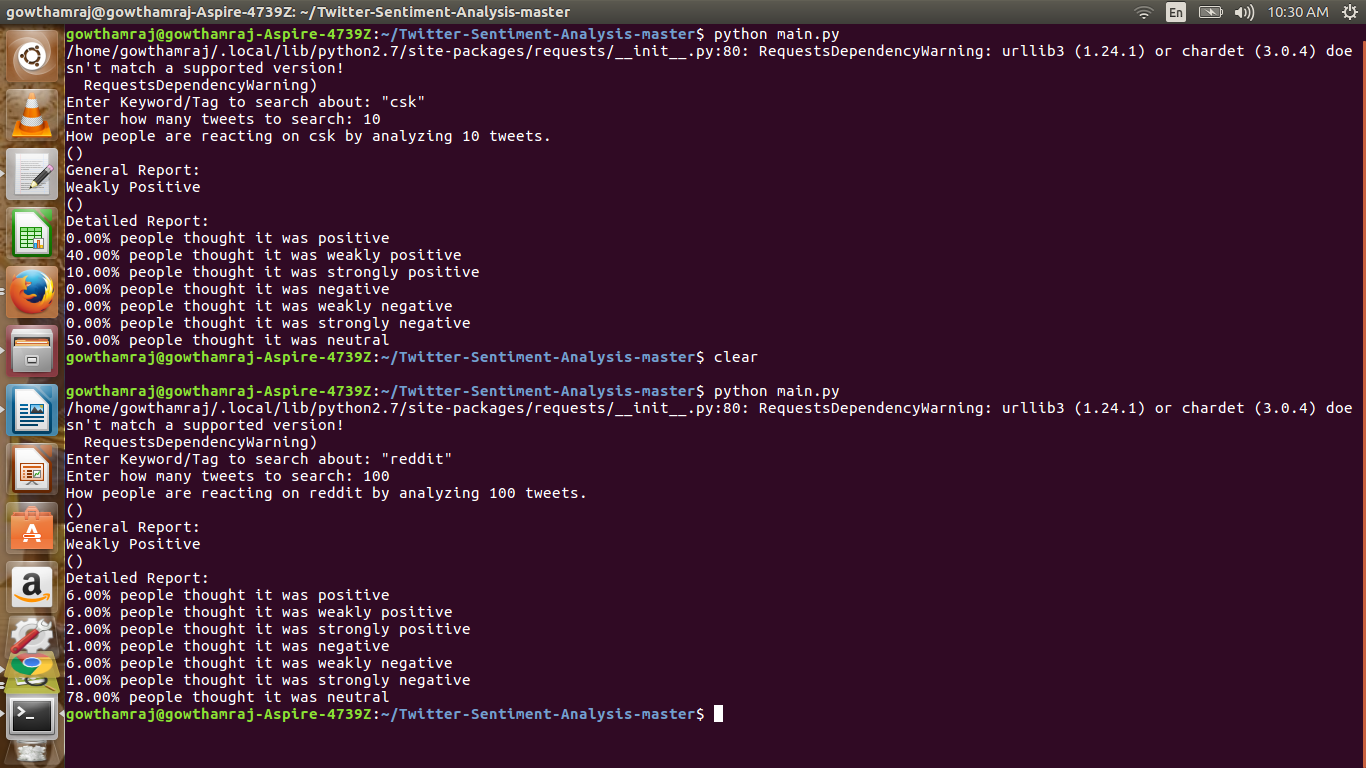


fig:report generation

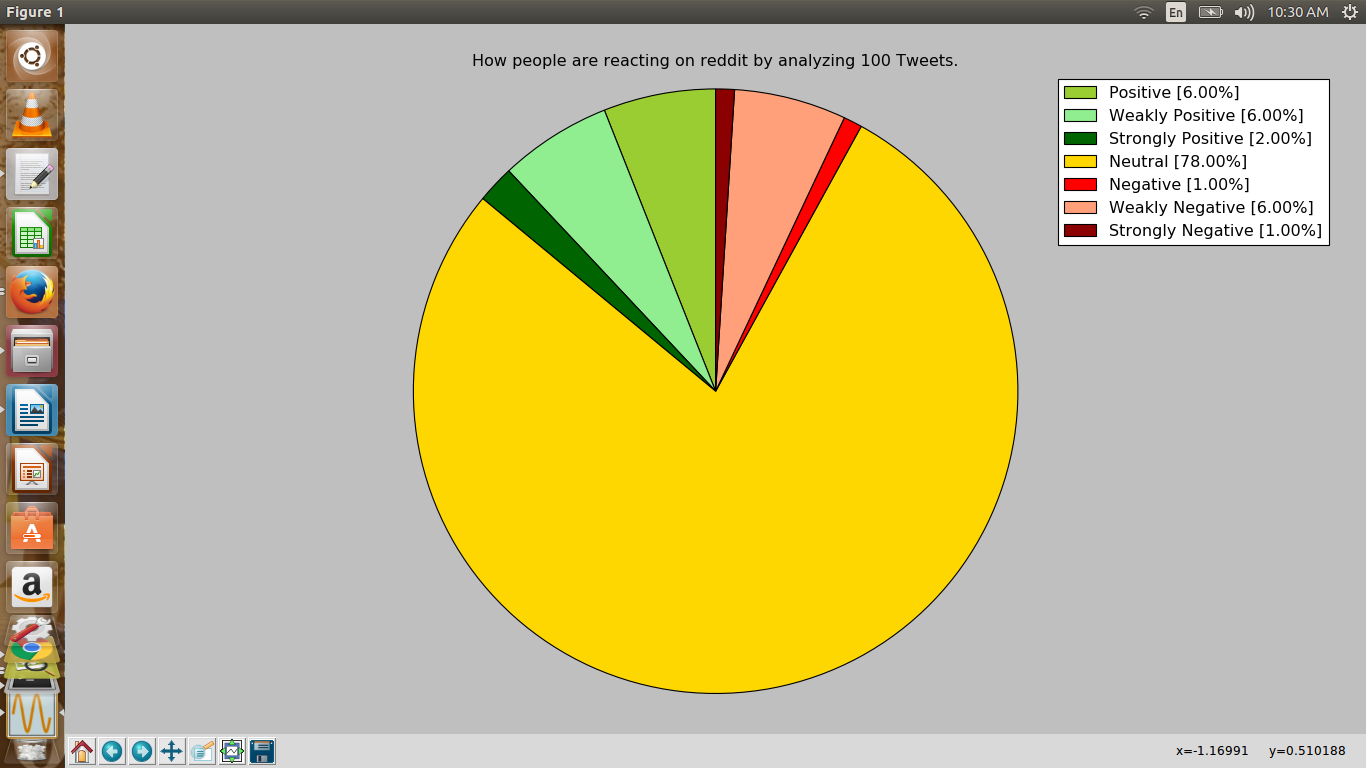


fig : visual representation of our analysis

Conclusion:

Twitter sentiment analysis is developed to analyze customers perspectives toward

the critical to success in the marketplace. The program is using a machine-based

learning approach which is more accurate for analyzing a sentiment; together with

natural language processing techniques will be used.

As a result, program will be categorized sentiment into positive and negative,

which is represented in a pie chart and html page Although, the program has been

planned to be developed as a web application, due to limitation of Django which

can only work on Linux server or LAMP. Thus, it cannot be realized. Therefore,

further enhancement of this element is recommended in future study.