CERTIFICATE

It is certified that the work contained in this thesis entitled "**Evaluation of product rating using data mining",** is by Priti Kumari (Roll no-1170449005) for the award of Master of Technology (M tech) from Babu Banarasi Das university, Lucknow has been carried out under my supervision and this work has not been submitted elsewhere for a degree.

Signature

Mrs. Upasana Dugal
(Assistant Professor)
Babu Banarasi Das University, Lucknow

ABSTRACT

In now days, suppositions and surveys open to us are a standout amongst the most basic factors in planning our perspectives and impacting the accomplishment of a brand, item or administration. With the coming and development of online life on the planet, partners regularly take to communicating their conclusions on well-known web based life, in particular twitter. While Twitter information is to a great degree educational, it introduces a test for examination in view of its humongous and disordered nature. This paper is an intensive push to plunge into the novel area of performing slant investigation of individuals' sentiments with respect to top universities in India. Other than taking extra preprocessing measures like the development of net language and expulsion of copy tweets.

ACKNOWLEDGMENTS

I take this opportunity to express a deep sense of gratitude towards my guide **Ms.Upasana Dugal,** for providing excellent guidance, encouragement, support and inspiration throughout the project work. Without her invaluable guidance, this work would never have been a success. I would also like to thank all my classmates for their valuable suggestions and helpful discussion specially in making report.

Priti kumari

BBDU, Lucknow

May, 2019

TABLE OF CONTENT

LIST OF FIGURE

LIST OF ABBREVIATION

JVM	Java Virtual Machine
VOC	voice of customer
ACK	Acknowledgement
SRS	Software Requirement Specification
VOC	voice of client
JDBC	Java Database Connectivity
GUI	graphical user interface

CHAPTER 1

INTRODUCTION

Online life has caught the consideration of the whole world as it is roaring quick in sending musings over the globe, easy to use and free of cost requiring just a working web association. Individuals are widely utilizing this stage to share their contemplations boisterous and clear. Twitter is one such understood smaller scale blogging website getting around 500 million tweets for every day .Each client has a day by day point of confinement of 2,400 tweets and 140 characters for every tweet. Twitter clients post (or 'tweet') each day about different subjects like items, administrations, everyday exercises, places, identities and so forth. Consequently, Twitter information is of Great pertinent as it can be utilized as a part of different situations where organizations or brands can use an immediate association with every one of their customer or client and in this way, enhance their item. Consider a not fulfilled costumer of a media transmission organization voicing out his/her grievances about a specific arrangement he/she is bought in to. Twitter additionally fills in as an enormous stage for clients to know progressively and get immediate remarks about an item or an administration in which they are intrigued. Sentiments and audits as tweets from clients, potential clients and pundits can undoubtedly impact the picture and subsequently, request of an item/benefit being given by an organization. Subsequently, regardless of whether the partner's Opinion is certain negative about their offering turns into an essential and squeezing question for the association to ask and screen.

According to fig. 1, roughly 34,582,000 out of an estimated176, 805, 000 of the 18-23 year old age group in India receive higher education which equates to about 19.56% of the age group [4]. Many reputed government and private colleges in India aim towards providing a class education to their students and follow different ideologies, pedagogies and examination procedures. It becomes highly important for the interested student to evaluate the choices available to him/her in selecting college that not only furnishes the student with the desired academic or professional prowess but also equips him/her with the right kind of learning tools according to his/her capabilities. Three of the premier colleges in India, namely the All India Institute of Medical Sciences (A.I.I.M.S.), the Indian Institute of

Technology (I.I.T.) and the National Institute of Technology(N.I.T.) [5] have been analyzed to find the user's sentiment pertaining to the perception of these colleges and the magnitude of these opinions.

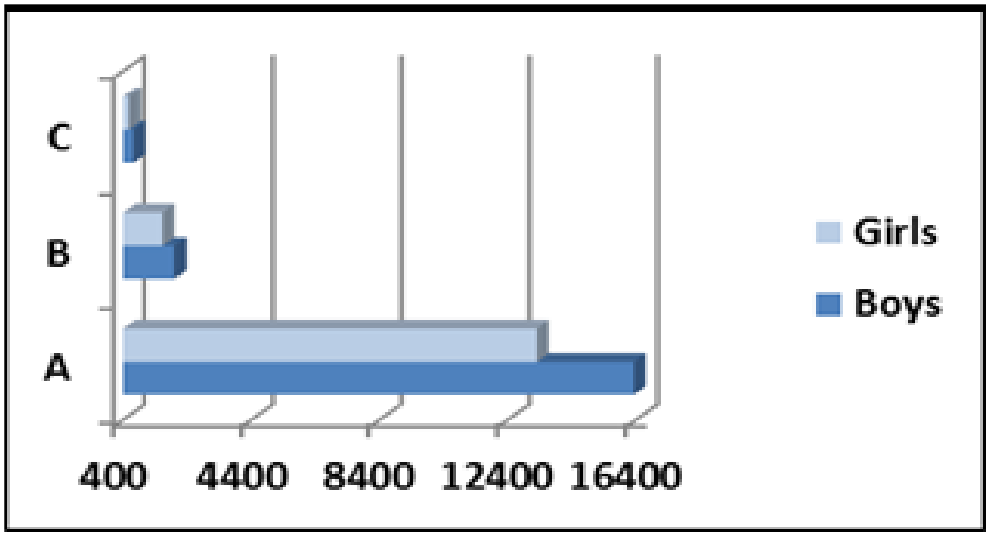


Fig. 1 Plot showing the enrollment of boys and girls of the 18-23 year old age group in Higher Education programs (based on the report compiled by the Ministry of Human Resource and Development in India). Here, A: Engineering, B: Medical , C:Architecture.

Sentiment Analysis has been of avid interest to researcher. A lot of work has been put into it and there is a vast IEEE Transaction on Computational Techniques in Information and Communication Technologies (March, 2016) domain of its applications. A number of studies focus upon the popularity and reviews of products and services offered by different organizations. Arora, Li and Neville used Lexicon based Sentiment analysis on various smart phone brands to judge their popularity and reviews in the range of sentiment scores from -6 to 6 [6]. Similarly, Choi, Lee, Park, Na and Choused sentiment analysis for laundry washers and televisions[7].Researchers have also been working upon prediction of accuracy of tested dataset using Machine Learning algorithms. Kanakaraj and Guddeti used Natural Language Processing

Techniques for sentiment analysis and compared Machine Learning Methods and Ensemble Methods to improve on the accuracy of the classification [8]. Bahrainian and Dengelcompared different supervised, unsupervised methods along with their hybrid

method (combining supervised, supervised methods) which outperformed other methods [9].Pak and Paroubek performed Sentiment Analysis using formulas of Entropy and Salience and also implemented Naïve Bayes and SVM [10]. Shahheidari, Dong and Bin Daud used a Naïve Bayes classifier for classification and tested it for news, finance, job, movies and sports taking into consideration data mining on the basis of two emoticons and [11]. Neethu M. S. and Rajasree R used twitter posts on electronic products, compared the accuracy between different machine learning algorithms and further improved the accuracy by replacing repeated characters with two occurrences, including a slang dictionary and taking emoticons into consideration [12].In addition, the area of neural networks has been investigated for performing sentiment analysis on benchmark datasets consisting of online product reviews. Bespalov, Bai, Qi and Shokoufandeh carried out binary classification on Amazon and Trip Advisor datasets using a Perception classifier and obtained one of the lowest error rates among their experiments of 7.59 and 7.37 on the two datasets respectively[13]. Jotheeswaran and Koteeswaran performed binary classification on the IMDB dataset by employing a Multi-layer Perception Neural Network and using Decision Tree-based Feature Ranking for feature extraction and a hybrid algorithm(based on Differential Evolution and Genetic Algorithm) for weight training, thereby obtaining a maximum classification accuracy of 83.25% [14]. Socher et al introduced a Semantic Treebank and a Recursive Neural Tensor Network which improves state of the art accuracy on binary classification from80% to 85.4% on the movie dataset introduced by Pang and Lee [15]. Santos and Gatti developed a deep convolution neural network and obtained an accuracy of 85.7% and 86.4%on the aforementioned Stanford Sentiment Treebank and Stanford Twitter Sentiment Corpus (which is bounded by its classification based on emoticons) respectively [16].

1.1 Problem Statement

The issue in conclusion examination is characterizing the extremity of a given content at the archive sentence or highlight/perspective level. Regardless of whether the communicated supposition in a record, a sentence or an element include/perspective is certain, negative, or impartial.

1.2 Objective

The objective of this proposed system is to give sentiment classification for dynamic tweets. The proposed system accuracy and efficiency is more compare to existing system. It should use dynamic tweets from tweeter for sentiment analysis and providing analysis of every single tweet.

1.2 EXISTING SYSTEM

In the existing system based on the initial expansion of the words they going to give sentiment process based on topics they adopted. In existing system excel sheet are used to analysis data. Initially in the existing system iteration process is done (for example 1st iteration they are going to take 100 tweets, within that 100 tweets which words are coming with more positive or more negative count that words will be added as positive or negative before 2nd iteration). It is not provided analysis based on every single statement. Here accuracy is less because after iteration immediately we considering positive or negative sentiment without considering left out words in tweets.

1.3 DISADVANTAGES OF EXISTING SYSTEM

- Time consuming
- Accuracy is less

1.4 PROPOSED SYSTEM

We are going to propose the system for detecting sentiment for dynamic tweets based on the threshold concept. Based on the threshold value we are going to achieve accuracy in the project. In the proposed system initial expansion is done based on the topic selected. Based on the topic every word in the particular tweet, sentiment type of word is checked. Finally positive, negative or neutral count is incremented. The left out word which is not in any sentiment type ,that word sentiment is decided based on the positive ,negative and neutral count in that particular tweet. If positive count is more than negative and neutral them it will be considered positive sentiment only. Finally if left out word sentiment crosses threshold value, then that word sentiment is permanently considered. In proposed system accuracy and efficiency is more compare to existing system. Proposed system using dynamic tweets from tweeter for sentiment analysis and providing analysis of every single tweet.

1.5 ADVANTAGES OF PROPOSED SYSTEM

- Accuracy
- Efficiency
- Analysis of every single tweet.
- Using dynamic tweets from tweeter

1.6 Proposed Algorithm

- **Preprocessing**

In this algorithm, the tweets which are foreign made to database from the twitter API, these tweets comprise of pointless words, whitespaces, hyperlinks and unique characters. First we have to do separating process by evacuating every single superfluous word, whitespaces, hyperlinks and extraordinary characters.

In this study, twitter data concerning three of the top colleges in India was obtained in JSON format for the duration of a month from 19 June, 2018 to 19 July, 2018. Unique

tweets referring to A.I.I.M.S., I.I.T. and N.I.T. were extracted in order to reduce the bias of user opinions, eliminate redundant data and minimize the frequency of tweets which may be spam or fake reviews. The tweets also provide information about the user, location, time-zone et cetera. In order to segregate the user opinion from user information, preprocessing was performed on the tweets. Removal of URLs, repeated letters in sequence which occurred more than twice with two of the same letter, ASCII escape sequences for Unicode characters, uninformative symbols and some but not all punctuations from the tweets was performed in order to sustain emoticons in the tweet. Expansion of SMS lingo, emoticons and abbreviations inner speak has been performed in order to include user opinions fitted rigidly under the constraint of 140 characters by referencing a slang dictionary which contains roughly 5,200slang words and incorporates about 270 emoticons. The preprocessing steps aim to begin the feature extraction process and start extracting bags of words from the samples. One of the main focus is to reduce the final amount of features extracted. Indeed, features reduction is important in order to improve the accuracy of the prediction for both topic modeling and sentiment analysis. Features are used to represent the samples, and the more the algorithm will be trained for a specific feature the more accurate the results will be. Hence, if two features are similar it is convenient to combine them as one unique feature. Moreover, if a feature is not relevant for the analysis, it can be removed from the bag of words.

- Lower uppercase letters: The first step in the preprocessing is to go through all the data and change every uppercase letter to their corresponding lowercase letter. When processing a word, the analysis will be case sensitive and the program will consider “data” and “Data” as two totally different words. It is important that, these two words are considered as the same features. Otherwise, the algorithms will affect sentiments which may differ to these two words. For example, on these three sentences: “data are good”, “Awesome data”, and “Bad Data”. The first and second sentences both contain “data” and are positive, the third sentence contains “Data” and is negative. The algorithm will guess that sentences containing “data” are more likely to be positive and those containing “Data” negative. If the uppercases had been removed the algorithm would have been able to guess that the fact that the

sentence contains “data” is not very relevant to detect whether or the sentence is positive. This preprocessing step is even more important since the data are retrieved from Twitter. Social media users are often writing in uppercase even if it is not required, thus this preprocessing step will have a better impact on social media data than other “classical” data.

- Remove URLs and user references: Twitter allows user to include hash tags, user references and URLs in their messages. In most cases, user references and URLs are not relevant for analyzing the content of a text. Therefore, this preprocessing step relies on regular expression to find and replace every URLs by “URL” and user reference by “AT_USER”, this allows to reduce the total amount of features extracted from the corpus [2]. The hash tags are not removed since they often contain a word which is relevant for the analysis, and the “#” characters will be removed during the tokenization process. Remove digits: Digits are not relevant for analyzing the data, so they can be removed from the sentences. Furthermore, in some cases digits will be mixed with words, removing them may allow to associate two features which may have been considered different by the algorithm otherwise. For example, some data may contain “I phone”, when other will contain “iphone7”. The tokenization process, which will be introduced later.

- Remove stop words: In natural language processing, stop words are often removed from the sample. These stop words are words which are commonly used in a language, and are not relevant for several natural language processing methods such as topic modeling and sentiment analysis [10]. Removing these words allows to reduce the amount of features extracted from the samples.

- **Self-Learning and word standardization System**

In this algorithm, first we have to instate the word reference (first emphasis dictionary).In the lexicon for the most part we have to introduce the positive, negative nonpartisan and things. Every single huge datum and information mining ventures in view of the prepared

information, without prepared information (introduction of words).So instatement of the prepared information is vital. In the self-learning framework, we are doing word institutionalization, here we are not considering past, present and future status of the words, just we are thinking about the word.

- **Sentiment Analysis**

In this algorithm, preprocessed tweets are brought from the database one by one. In the first place we require check one by one watchword whether that catchphrase is thing are not, if thing we will expel it from the specific tweet. After that the rest of the watchwords checked with assessment compose, regardless of whether that catchphrases are certain opinion or negative conclusion or impartial feeling. The rest of the watchwords in the tweet which does not has a place with any of the supposition will be relegated transitory conclusion in light of the more check of positive, negative and impartial. In the second cycle if the reaming word crosses the limit of positive, negative or nonpartisan, that watchword forever included as development in the lexicon. At long last opinion of the tweet is recognized in light of the positive, negative and impartial words in the specific tweet.

1.6 Thesis Organization

The thesis scope is to demonstrate the main issues that are related to the Tweeter Sentiment analysis, and plotting graph of its analysis, including the application, main duties, etc.

Chapter two demonstrates main theoretical topics and illustrations. It is a summary for the material related to clustering and protocols of the tweeter sentiment analysis, in addition to major structure.

The Software requirement analysis work is shown in Chapter three in details. It illustrates the main steps of work, diagram of development process and program running, also, the program flow in details. Chapter four shows the system design details, with different measurements.

CHAPTER 2

LITERATURE SURVEY

Based on broad writing review identified with the Twitter assessment mining of Top Colleges in India has been thought about in this part.

Conclusion Analysis has been of ardent enthusiasm to analysts recently. A ton of work has been put into it and there is an immense prevalence and surveys of items and administrations offered by various associations. Arora, Li and Neville utilized Lexicon construct Sentiment investigation in light of different advanced cell brands to judge their fame and audits in the scope of estimation scores from - 6 to 6 [6]. So also, Choi, Lee, Park, Na and Cho utilized conclusion examination for clothing washers and TVs [7]. Specialists have additionally been endless supply of precision of tried dataset utilizing Machine Learning calculations. Kanakaraj and Guddeti utilized Natural Language Processing Techniques for assumption investigation and thought about Machine Learning Methods and Ensemble Methods to enhance the exactness of the characterization [8]. Bahrainian and Dengel analyzed diverse directed, unsupervised techniques alongside their half breed strategy (joining regulated and unsupervised strategies) which beat different strategies [9]. Pak and Paroubek performed Sentiment Analysis utilizing equations of Entropy and Salience and furthermore actualized Naïve Bayes and SVM [10]. Shahheidari, Dong and Bin Daud utilized a Naïve Bayes classifier for arrangement and tried it for news, back, occupation, motion pictures and games contemplating information mining based on two emojis and [11]. Neethu M. S. what's more, Rajasree R utilized twitter posts on electronic items, thought about the precision between various machine learning calculations and further enhanced the exactness by supplanting rehashed characters with two events, including a slang word reference and mulling over emojis [12]. Likewise, the territory of neural systems has been researched for performing assessment investigation on benchmark datasets comprising of online item audits. Bespalov, Bai, Qi and Shokoufandeh completed double arrangement on Amazon and TripAdvisor datasets utilizing a Perceptron classifier and got one of the most reduced blunder rates among their examinations of 7.59 and 7.37 on the two datasets separately [13]. Jotheeswaran and Koteeswaran performed double characterization on the IMDB dataset by utilizing a Multi-layer Perception Neural

Network and utilizing Decision Tree-based Feature Ranking for highlight extraction and a half and half calculation (in view of Differential Evolution and Genetic Algorithm) for weight preparing, in this way acquiring a most extreme arrangement precision of 83.25% [14]. Socher et al presented a Semantic Treebank and a Recursive Neural Tensor Network which enhances cutting edge precision on double arrangement from 80% to 85.4% on the film dataset presented by Pang and Lee [15]. Santos and Gatti built up a profound convolutional neural system and got an exactness of 85.7% and 86.4% on the previously mentioned Stanford Sentiment Treebank and Stanford Twitter Sentiment Corpus (which is limited by its order in light of emojis) separately [16].

King R. A., Racherla P. and Bush V. D., 2014 Electronic word-of mouth (eWOM) has attracted considerable interest from researchers in the past decade. Although the extant research has helped us to develop a good understanding of a number of the issues pertaining to eWOM, several research and managerial questions remain. Furthermore, no attempt has been made to consolidate and synthesize this stream of research. With consumers' increasing reliance on online retailing and information seeking, as well as the continued growth of social media, the importance of eWOM cannot be overstated. Based on a systematic review of 190 studies, we conduct a multi-dimensional analysis of eWOM communication. We present the key issues in current and emerging literature and propose important questions for future research.

Arora D., Li K.F. and Neville S.W., 2015 Intoday‟s era, which is symbolic of „information overload‟, we are living in a time where knowledge needs to be mined from information which is available in abundance, one such instance being mining sentiments from user textual reviews which plays a decisive role in consumer’s decision making processes. Today’s customers are intelligent enough and will cross-check before purchasing any product and the best place for that are online reviews which are ubiquitous. In today’s competitive market, analyzing user needs and sentiments is of uttermost importance as one bad review can make or break the profits of a brand since Word of Mouth plays a significant role in building a customer base for a brand or product. An

approach for consumer s sentiment analysis has been proposed, which will mine user's sentiments and thus pave a foundation for generating the popularity of the products which will lead to „personalized" results which are necessary in today s client-centric world. This in turn is helpful for brands wanting to expand the horizon of their user turnover and to devise better retailing strategies to sell their products. The approach is to convert these textual reviews into star ratings that will describe the „likeability" and popularity of the products.

Kanakaraj M., Guddeti R M.R., 2015 Mining opinions and analyzing sentiments from social network data help in various fields such as even prediction, analyzing overall mood of public on a particular social issue and so on. This paper involves analyzing the mood of the society on a particular news from Twitter posts. The key idea of the paper is to increase the accuracy of classification by including Natural Language Processing Techniques (NLP) especially semantics and Word Sense Disambiguation. The mined text information is subjected to Ensemble classification to analyze the sentiment. Ensemble classification involves combining the effect of various independent classifiers on a particular classification problem. Experiments conducted demonstrate that ensemble classifier outperforms traditional machine learning classifiers by 3-5%.

Neethu M. S. and Rajasree R., 2013 Sentiment analysis deals with identifying and classifying opinions or sentiments expressed in source text. Social media is generating a vast amount of sentiment rich data in the form of tweets, status updates, blog posts etc. Sentiment analysis of this user generated data is very useful in knowing the opinion of the crowd. Twitter sentiment analysis is difficult compared to general sentiment analysis due to the presence of slang words and misspellings. The maximum limit of characters that are allowed in Twitter is 140. Knowledge base approach and Machine learning approach are the two strategies used for analyzing sentiments from the text. In this paper, we try to analyze the twitter posts about electronic products like mobiles, laptops etc using Machine Learning approach. By doing sentiment analysis in a specific domain, it is possible to identify the effect of domain information in sentiment classification. We present a new

feature vector for classifying the tweets as positive, negative and extract peoples' opinion about products.

Bespalov D., Bai B., Qi Y., and Shokoufandeh A, 2011we propose an efficient embedding for modeling higherorder (n-gram) phrases that projects the n-grams to low-dimensional latent semantic space, where a classification function can be defined. We utilize a deep neural network to build a unified discriminative framework that allows for estimating the parameters of the latent space as well as the classification function with a bias for the target classification task at hand. We apply the framework to large-scale sentimental classification task. We present comparative evaluation of the proposed method on two (large) benchmark data sets for online product reviews. The proposed method achieves superior performance in comparison to the state of the art.

Jotheeswaran J. and Koteeswaran S, 2015 ,Sentiment analysis plays a big role in brand and product positioning, consumer attitude detection, market research and customer relationship management. Essential part of information-gathering for market research is to find the opinion of people about the product. With availability and popularity of like online review sites and personal blogs, more chances and challenges arise as people now can, and do use information technologies to understand others opinions. In this paper, a Multi-Layer Perceptron (MLP) is used to classify the features extracted from the movie reviews. A Decision Tree-based Feature Ranking is proposed for feature selection. The ranking is based on Manhattan Hierarchical Cluster Criterion In the proposed feature selection; a decision tree induction selects relevant features. Decision tree induction constructs a tree structure with internal nodes denoting an attribute test with the branch representing test outcome and external node denotes class prediction. In this paper, a hybrid algorithm based on Differential Evolution (DE) and Genetic Algorithm (GA) for weight optimization algorithm to optimize MLPNN is proposed. IMDb dataset is used to evaluate the proposed method. Experimental results showed that the MLP with proposed feature selection improves the performance of MLP significantly by 3.96% to 6.56%. Classification accuracy of 81.25% was achieved when 70 or 90 features were selected.

Socher R., 2013 Semantic word spaces have been very useful but cannot express the meaning of longer phrases in a principled way. Further progress towards understanding compositionality in tasks such as sentiment detection requires richer supervised training and evaluation resources and more powerful models of composition. To remedy this, we introduce a Sentiment Treebank. It includes fine grained sentiment labels for 215,154 phrases in the parse trees of 11,855 sentences and presents new challenges for sentiment compositionality. To address them, we introduce the Recursive Neural Tensor Network. When trained on the new tree bank, this model outperforms all previous methods on several metrics. It pushes the state of the art in single sentence positive/negative classification from 80% up to 85.4%. The accuracy of predicting fine-grained sentiment labels for all phrases reaches 80.7%, an improvement of 9.7% over bag of features baselines. Lastly, it is the only model that can accurately capture the effects of negation and its scope at various tree levels for both positive and negative phrases.

Ng A.Y., Jordan M. I.,2002 We compare discriminative and generative learning as typified by logistic regression and naive Bayes. We show, contrary to a widely held belief that discriminative classifiers are almost always to be preferred, that there can often be two distinct regimes of performance as the training set size is increased, one in which each algorithm does better. This stems from the observation- which is borne out in repeated experiments- that while discriminative learning has lower asymptotic error, a generative classifier may also approach its (higher) asymptotic error much faster.

Koto F. and Adriani M, 2015, In this paper, investigations of Sentiment Analysis over well-known Social Media Twitter were done. As literatures show that some works related to Twitter Sentiment Analysis have been done and delivered interesting idea of features, but there is no a comparative study that shows the best features in performing Sentiment Analysis. In totalwe used 9 feature sets (41 attributes) that comprise punctuation, lexical, part of speech, emoticon, Senti Word lexicon, AFINN-lexicon, Opinion lexicon, Senti-Strength method, and Emotion lexicon. Feature analysis was done by conducting

supervised classification for each feature set sand continued with feature selection in subjectivity and polarity domain. By using four different datasets, the results reveal that AFINN lexicon and Senti-Strength method are the best current approaches to perform Twitter Sentiment Analysis.

dos Santos C. N. and Gatti M, 200 Sentiment analysis of short texts such as single sentences and Twitter messages is challenging because of the limited contextual information that they normally contain. Effectively solving this task requires strategies that combine the small text content with prior knowledge and use more than just bag-of-words. In this work we propose a new deep convolution neural network that exploits from character- to sentence-level information to perform sentiment analysis of short texts. We apply our approach for two corpora of two different domains: the Stanford Sentiment Treebank (SSTb), which contains sentences from movie reviews; and the Stanford Twitter Sentiment corpus (STS), which contains Twitter messages. For the SSTb corpus, our approach achieves state-of-the-art results for single sentence sentiment prediction in both binary positive/negative classification, with 85.7% accuracy, and fine-grained classification, with 48.3% accuracy. For the STS corpus, our approach achieves a sentiment prediction accuracy of 86.4%.

Salazar D. A., Velez J. I., Salazar J. C, 2012, The classification of individuals is a common problem in applied statistics. If X is a data set corresponding to a sample from an specific population in which observations belong to g different categories, the goal of classification methods is to determine to which of them a new observation will belong to. When g = 2, logistic regression (LR) is one of the most widely used classification methods. More recently, Support Vector Machines (SVM) has become an important alternative. In this paper, the fundamentals of LR and SVM are described, and the question of which one is better to discriminate is addressed using statistical simulation. An application with real data from a microarray experiment is presented as illustration.

Rajendran S., Kalpana B, 2011 Support Vector machine (SVM) has become an optimistic method for data mining and machine learning. The exploit of SVM gave rise to

the development of a new class of theoretically refined learning machines, which uses a central concept of kernels and the associated reproducing kernel Hilbert space. The performance of SVM largely depends on the kernel. However, there is no premise about how to choose a good kernel function for a particular domain. This paper focuses in this issue i.e. the choice of the Kernel Function is studied empirically and optimal results are achieved for binary class SVMs. The performance of the Binary class SVM is illustrated by extensive experimental results. The experimental results of the datasets show that RBF Kernel or any other kernels is not always the best choice to achieve high generalization of classifier although it is often the default choice.

Chi-Hwan Choi, Jeong-Eun Lee, IJSH, 2013 Advances in Internet and Smartphones are producing colossal measure of clients' supposition (or VOC: voice of client) in the online world. VOC is exceptionally valuable for showcasing or item outline offices to see the clients' assessment for the items. In this paper, VOC in online audit destinations has been gathered, put away and examined keeping in mind the end goal to discover what sort of elements can influence the client inclination for the gadgets items. We have executed a VOC investigation framework that incorporates ETL (concentrate, change, and stacking), regular dialect handling, positive/negative examination, and result perception. Affectability modifiers and equivalent words are gathered and put away as a database for more right assessment examination. Genuine test for VOC information with 39 Giga bytes from surely understood survey destinations in Korea demonstrates that the accuracy for the characterization of the positive/negative sentences and the reasons for the positive/negative are 73% and 70%, separately. The assessment has been done in light of the manual examination for the investigation result.

SaeidehShahheidari, Hai Dong, IEEE, 2013 Nowadays, micro blogging turns out to be progressively mainstream among individuals as individuals can impart their insights about each part of their life through this stage. Sentiment mining is a strategy to remove information from the assessments that individuals share in web gatherings, sites, talk gatherings, and remark boxes. Moreover, sentiment mining utilizes content mining and

normal dialect preparing strategies to influence PC to comprehend the outflow of feelings. Be that as it may, its fundamental concern is to remove nostalgic and enthusiastic articulations from unstructured content. Recognizing the best technique for grouping is a basic errand for slant examination. A large number of the methodologies depend on database for feeling investigation. In this way, the fundamental goal of this paper is to plan a straightforward however speedier conclusion classifier
by using the Twitter corpus as the database for putting away notion designs. Our second target is to examine in which area assumption can be best grouped with this new classifier. By executing this approach, we can diminish the improvement time of building a precise conclusion classifier for a chose space. To understand these two goals, we made utilization of the Naive Bayes calculation to group the extremity of individuals' sentiment about various subjects on Twitter. It is normal that the time and many-sided quality utilized for characterizing the nostalgia of tweets would be significantly decreased by methods for this approach.

Bo Pang and Lillian Lee a vital piece of our data gathering conduct has dependably been to discover what other individuals think. With the developing accessibility and fame of conclusion rich assets, for example, online audit locales and individual sites, new openings and difficulties emerge as individuals presently can, and do, effectively utilize data advancements to search out and comprehend the conclusions of others. The sudden ejection of action in the zone of assessment mining and assumption investigation, which manages the computational treatment of feeling, notion, and subjectivity in content, has in this manner happened in any event to some degree as an immediate reaction to the surge of enthusiasm for new frameworks that arrangement specifically with conclusions as a top of the line protest.

This study covers procedures and methodologies that guarantee to straightforwardly empower sentiment arranged data looking for frameworks. Our emphasis is on techniques that try to address the new difficulties raised by notion mindful applications, when contrasted with those that are now present in more conventional certainty based examination. We incorporate material on synopsis of evaluative content and on more

extensive issues with respect to protection, control, and monetary effect that the advancement of conclusion arranged data get to administrations offers ascend to. To encourage future work, a discourse of accessible assets, benchmark datasets, and assessment battles is additionally given.

Alexander Pak, Patrick Paroubek Micro blogging today has turned into an extremely well known specialized device among Internet clients. A huge number of clients share feelings on various parts of life consistently. Accordingly smaller scale blogging sites are rich wellsprings of information for assessment mining and opinion investigation. Since smaller scale blogging has showed up moderately as of late, there are a couple of research works that were committed to this subject. In our paper, we center around utilizing Twitter, the most prevalent smaller scale blogging stage, for the assignment of conclusion investigation. We demonstrate to naturally gather a corpus for conclusion investigation and assessment mining purposes. We perform etymological investigation of the gathered corpus and clarify found marvels. Utilizing the corpus, we construct an estimation classifier that can decide positive, negative and unbiased assessments for a report. Test assessments demonstrate that our proposed procedures are productive and perform superior to already proposed techniques. In our exploration, we worked with English; be that as it may, the proposed method can be utilized with some other dialect.

Seyed-Ali Bahrainian, Andreas Dengel, IEEE, 2013 Sentiment Analysis (SA) and rundown has as of late turned into the focal point of numerous specialists, since examination of online content is valuable and requested in a wide range of uses. One such application is item based opinion outline of multi-records with the motivation behind illuminating clients about upsides and downsides of different items. This paper acquaints a novel arrangement with target-situated (i.e. perspective based) supposition outline and SA of short casual writings with a primary spotlight on Twitter posts known as "tweets". We think about various calculations and techniques for SA extremity recognition and opinion synopsis. We demonstrate that our half breed extremity location framework not just beats the unigram cutting edge gauge, yet in addition could be favorable position over different

techniques when utilized as a piece of a notion outline framework. Also, we represent that our SA and rundown framework shows a superior with different helpful functionalities and highlights.

Conclusion

This part generally discusses the papers that are suggested while influencing this proposal to report. Each one of these papers give information related to learning of total lead, their present courses of action, and systems used moreover their central focuses and imperatives.

CHAPTER 3

SOFTWARE REQUIREMENT SPECIFICATION

This part depicts about the prerequisites. It determines the equipment and programming prerequisite that are needed for software to keeping in mind the end goal, to run the application appropriately. The Software Requirement Specification (SRS) is clarified in point of interest, which incorporates outline of this exposition and additionally the functional and non-practical necessity of this thesis.

3.1 General Description

The reason behind the framework prerequisites and determination record is to depict the assets and administration of those assets utilized as a part of the configuration of the Public Key Cryptosystem for information partaking in distributed storage. This framework necessity and particular will likewise give insights with respect to the utilitarian and non-useful prerequisites of the venture.

3.1.1 Users Perspective

The Characteristic of this task work is to give information adaptability security while sharing information through cloud. It gives a proficient approach to share information through cloud.

3.2 Non Functional Requirement

Non-utilitarian necessities are the prerequisites which are not straightforwardly having a place with the specific capacity gave by the framework. This gives the criteria that can be utilized to finish up the operation of a framework rather than particular practices.

This can be utilized to relate the rising structure properties, for instance, immovable quality, response time and store inhabitancies. Here again they ought to portray objectives on the system, for instance, the capacity of the data yield devices and data representation used as a piece of structure interfaces. In all probability all non-helpful essentials can be relating to the system as whole rather than to individual structure highlights. This suggests

they are every now and again essential appear differently in relation to the individual commonsense necessities. Non utilitarian necessity gets through the client needs, in view of spending plan limitations, hierarchical approaches, and the requirement for interoperability with other programming and equipment frameworks.

The going with non-valuable requirements are meriting thought.

- Security: The framework ought to permit a secured correspondence between information proprietor and beneficiary.
- Reliability: The system should be trustworthy and not corrupt the execution of the present structure and should not to provoke the hanging of the structure.

3.3 System Requirement

3.3.1 Hardware Requirement

- **Processor** **:** intel
- **Keyboard** **:** 104 Keys
- **Floppy Drive** **:** 1.44 MB MHz Pentium III
- **RAM** **:** 128 MB
- **Hard Disk** **:** 10 GB
- **Monitor** **:** 14” VGA COLOR
- **Mouse** **:** Logitech Serial Mouse
- **Disk Space** **:** 1 GB

3.3.2 Software Requirements

- **Operating System** **:** Win 2000/ XP
- **Server** **:** Apache Tomcat
- **Technologies used** **:** Java, Servlets, JSP, JDBC
- **JDK** **:** Version 1.7
- **Database** **:** My SQL 5.0

3.4 Feasibility Study

Believability is the determination of paying little respect to whether an undertaking justifies action. The framework followed in building their strength is called acceptability Study, these kind of study if a task could and ought to be taken.

Three key thoughts included in the likelihood examination are:

- Technical Feasibility
- Economic Feasibility
- Operational Feasibility

3.4.1 Technical Feasibility

Here it is considered with determining hardware and programming, this will effective fulfill the client necessity the specialized requires of the framework should shift significantly yet may incorporate

- The office to create yields in a specified time.
- Reaction time under particular states.
- Capacity to deal with a particular segment of exchange at a specific pace.

3.4.2 Economic Feasibility

Budgetary examination is the often used system for assessing the feasibility of a projected structure. This is more usually acknowledged as cost/favourable position examination. The method is to center the focal points and trusts are typical casing a projected structure and a difference them and charges. These points of interest surpass costs; a choice is engaged to diagram and realize the system will must be prepared if there is to have a probability of being embraced. There is a consistent attempt that upgrades in exactness at all time of the system life cycle.

3.4.3 Operational Feasibility

It is for the most part identified with human association and supporting angles. The focuses are considered:

- ❖ What alterations will be carried through the framework?
- ❖ What authoritative shapes are dispersed?
- ❖ What new aptitudes will be needed?
- ❖ Do the current framework employee's individuals have these aptitudes?
- ❖ If not, would they be able to be prepared over the span of time?

3.5 Resource Requirement

3.5.1 Java

Java is a stage autonomous programming dialect. It is outline to be basic and convenient crosswise over diverse stages.

The java programming vernacular is an unusual state tongue that can be portrayed by most of the going with in vogue expressions:

- ➢ Object oriented
- ➢ Simple
- ➢ Architecture neutral
- ➢ Portable
- ➢ Robust
- ➢ Dynamic
- ➢ Secure

The Java API is a broad social affair of moment programming fragments that give various profitable limits, for instance, graphical user interface (GUI) contraptions. The Java API is accumulated into collections of correlated classes and interfaces; these collections are recognized as packs.

Java stage gives you the accompanying elements:

- ➢ **The essentials:** Items, strings, strings, numbers, info, yield, information structures, framework properties, date, time et cetera.

- **Applets:** The arrangement of traditions utilized by applets.
- **Networking:** URLs, TCP, UDP attachments, and IP addresses.
- **Internationalization:** Help for composing projects that could be restricted for clients around the world. Projects can naturally adjust to particular local people and be shown in the suitable dialect.
- **Security:** Mutually low level and abnormal state, together with electronic marks, open and private key administration, right of entry control and authentications.
- **Software components:** Recognized as JavaBeans, could connect to existing parts structural designs.
- **Object Serialization:** Let's Permits lightweight tirelessness and correspondence by means of Remote Method Invocation (RMI).
- **Java Database Connectivity (JDBC):** Give consistent entree to an extensive variety of social databases.

Advantage of java technology:

- **Get started quickly:** In spite of the fact that the java programming dialect is an intense article arranged dialect, it is anything but difficult to learn, particularly for software engineers effectively acquainted using C or C++.
- **Write less code:** Examinations of undertaking estimation advise that a framework built in the java language tongue shall be 4 times more diminutive compare to similar program in C++.
- **Write better code:** Java dialect energizes great coding rehearsals and its trash gathering serves to evade memory spills. Its item introduction, its javaBeans segment building design and its far reaching, effectively extendible API let's to use again other individuals tried code and present less bugs.
- **Develop programs more quickly:** Headway time can be as much as twice as speedy against making the similar program in C++.
- **Write once, run anywhere:** Since 100% immaculate java projects are ordered into system autonomous byte codes, run reliably on whichever java stage.

- **Distribute software more easily:** Update applets smoothly from a middle server. The applets misuse the segment of allocating new classes could be stacked "on the fly", without recompiling the whole system.

3.5.2 Java Server Page

Java Server Pages development allows you to put scraps of servlet coding particularly into a substance based record. A java server page is a substance based record that holds two sorts of substance: static format data, which should be imparted in several substance based association, for instance, XML, WML, HTML and Java server page segments, which choose how the page fabricates component content.

Java Server Page (JSP): An extensible Web innovation that make use of layout information, custom components, scripting dialects, and server side Java programming language articles to homecoming element substance to a customer. As indicated by JSP model1 we can build up the application as:

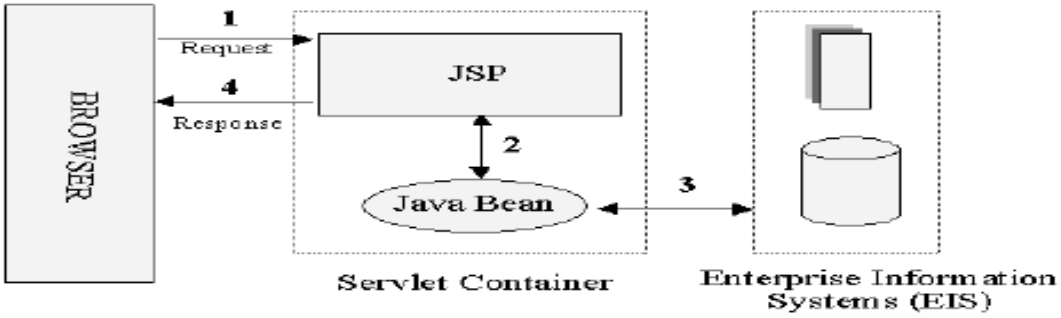


Figure 3.5.2.1 Architecture of jsp model 1

Commonly the layout information is XML/HTML components, and as a rule the customer is a Web programs. According to above replica the presentation method of reasoning must be executed in java server page and the business justification must be realized as a component of Java bean and this model assist us in disconnecting the appearance and business basis.

For sweeping extent reaches out instead of using model1 it is perfect to use model2 Model View Controller. Struts structure is in light of model 2.JSP allows you to specific the dynamic bit of your pages with the static HTML. Here it just makes the steady HTML in the run of the mill way, using any Web page building mechanical assemblies it frequently

uses. Here again encase the dynamic code for the parts in exceptional labels, the greater part of which begin with "<%" and end with "%>". You ordinarily give your record a .jsp augmentation, and regularly introduce it in wherever you could put a typical Web page.

Despite the fact that what you compose frequently seems more like a customary HTML record compare to a servlet, in the background, the java server page just acquires change over to an average servlet, through the static HTML basically being printed to the yield stream joined with the servlet's organization method.

3.5.3 JavaScript

Java Script is a one of the object oriented; lightweight, script based computer programming language dialect tongue which was made by Netscape Communication Corporation. JavaScript holds the change equally client and server parts of application of Web program and the client side; this could be employ to form programs which are implemented by a Web program inside the association of a Web page on the server side, this should be employed to make a Web server programs which could deal with information set up together by a Web project and thereafter overhaul the program's showcase moreover.

Despite the fact that JavaScript underpins client and server Web programming, this lean toward JavaScript at customer side programming subsequent to a large portion of the programs bolsters this one. JavaScript is as simple as to study as HTML, and JavaScript explanations could be incorporated in HTML reports by encasing the announcements among a couple of scripting labels.

Here there are a couple of issues we can do with JavaScript

- Approve the substance of a structure and build computations.
- Include looking over or changing information to the Browser's status line.
- Invigorate pictures or pivot pictures that change when we shift the mouse above them.
- Identify the program being used and presentation distinctive substance for diverse programs.
- Identify introduced modules and inform the client if a module is needed.

- JavaScript can do a great deal more with JavaScript, including making whole application.

JavaScript and Java are altogether diverse dialects. A couple of the obvious contrasts are:

- Java applets are usually indicated for a situation inside the web chronicle; JavaScript could impact every bit of the Web report itself.
- JavaScript is most suitable to essential applications and inserting instinctive components to Web pages and Java could be used for amazingly multifaceted applications.

Here there are various distinctive complexities yet the fundamental object to review is that JavaScript and Java are autonomous lingos. They are together useful for unmistakable objects; really they could be used both to join their ideal circumstances.

3.5.4 JDBC

JDBC with a last goal to set a self-administering database benchmark API for java made database blend, or JDBC. This offers a nonspecific SQL database access portion that provides an expected interface to a mixed pack of RDBMS. This trustworthy interface is refined in the use of "unit" database framework units, or drivers. On the off chance that a database merchant wishes to have JDBC strengthen, he or she can allot to the driver to every stage that the database and java keep running on. To get a more expansive attestation of JDBC, sun builds up JDBC's system concerning ODBC. Java database integration gives uniform access to an extensive variety of social databases. MS Access databases are utilized for rapidly overhauling the store table.

Plan objectives for JDBC are as per the following:

- **SQL Level API**

The organizers experience that their essential objective was to portray a SQL interface for java. Yet not the most diminished database interface level achievable, this is an adequately low level for strange state devices and APIs to be made. Then again, it is at an adequately abnormal state for application programming architect to use it surely. Finishing this target looks into future mechanical assembly shippers to "make" JDBC code and to cover countless difficulties from the end customer.

- **SQL Conformance**

  SQL sentence structure changes as this movement from database shipper to database vender. With an end target to support a wide grouping of traders; JDBC will authorize every investigation articulation to be moved out through this to the basic database driver. This authorizes the integration unit to grip non-standard usefulness in approach that is appropriate for their clients.

- **JDBC must be implemented on top of basic database interfaces**

  The JDBC SQL API must "sit" on top of other fundamental SQL level APIs. This target grants JDBC to use active ODBC level drivers by the usage of an item interface. This interface could be making an elucidation of JDBC calls to ODBC and the other route around.

- **Give a java interface that is steady with whatever is left of the java framework**

  As of java's affirmation in the customer gathering thusly for, the makers feels that they should not to wander away from the present blueprint of the middle java structure.

- **Keep it simple**

  These objective likely shows up in all products outline objective postings. JDBC is no special case. Sun considered that the configuration of JDBC ought to be extremely basic, taking into account stand out strategy for finishing an undertaking for every component. Permitting copy usefulness just server to confound the client of the API.

- **Keep the common cases simple**

Since as a general rule, the typical SQL calls utilize the software engineers are straightforward INSERT's, SELECTs, UPDATE's and DELETE's these request should be anything but difficult to perform with JDBC. Be that as it may, more intricate SQL explanations ought to likewise be conceivable.

4. Tools:

The tools are used to trace the opinion or polarity from the user generated texts.

They are many tools present for data mining here we are using WEKA tool.

WEKA

WEKA is java based free and open source, software licensed and available for used on Linux, Mac OS X and window. It consists of machine learning algorithm for data mining, package tool for data pre-processing, classification, clustering, association, regression. It is idea choose for educational, research purpose as well as prototype.

Conclusion

This part gives subtle elements of the practical prerequisites, non-utilitarian necessities, asset prerequisites, equipment necessities, programming necessities and so on. Again the non-utilitarian prerequisites thus contain item necessities, authoritative prerequisites, client prerequisites, fundamental operational necessities and so on.

CHAPTER 4

SYSTEM ANALYSIS

4.1 Introduction to System Analysis

4.1.1 System

A system is an orderly group of interdependent components linked together according to a plan to achieve a specific objective. Its main characteristics are organization, interaction, interdependence, integration and a central objective.

4.1.2 System Analysis

System analysis and design are the application of the system approach to problem solving generally using computers. To reconstruct a system the analyst must consider its elements output and inputs, processors, controls feedback and environment.

4.2 Existing System

In the current framework in light of the underlying development of the words they going to give supposition process in light of points they received. At first in the current framework cycle process is done (for instance first emphasis they will take 100 tweets, inside that 100 tweets which words are accompanying more positive or more negative tally that words will be included as positive or negative before second iteration).Here precision is less in light of the fact that after cycle quickly we thinking about positive or negative slant without thinking about left out words in tweets.

4.3 DISADVANTAGES

- Time consuming
- Accuracy is less

4.4 Proposed System

We will propose the framework for recognizing notion for dynamic tweets in view of the edge idea. In view of the limit esteem we will accomplish exactness in the undertaking. In the proposed framework starting extension is done in view of the subject chose. In light of the point each word in the specific tweet, slant sort of word is checked. At last positive, negative or unbiased tally is augmented. The left out word which isn't in any slant compose ,that word assumption is chosen in view of the positive ,negative and unbiased tally in that specific tweet. In the event that positive tally is more than negative and unbiased them it will be viewed as positive estimation as it were. At long last if left out word opinion crosses edge esteem, at that point that word supposition is for all time considered.

4.5 Advantages of the Proposed System

- Accuracy
- Efficiency

4.6 Project Module Description

- **Tweets Import Module**

In this module, tweets are recovered from the twitter API powerfully in view of the school name input. To recover tweets from the twitter API account, first need to make twitter account in engineer's support. In the wake of making account we will get shoppers token key and access token key, with the assistance of produced keys, we will speak with twitter API to recover tweets. The recovered tweets are transported in into database.

- **Preprocessing Module**

In this module, the tweets which are foreign made to database from the twitter API, these tweets comprise of pointless words, whitespaces, hyperlinks and unique characters.First we have to do separating process by evacuating every single superfluous word, whitespaces, hyperlinks and extraordinary characters.

- **Self-Learning and word standardization System**

In this module, first we have to instate the word reference (first emphasis dictionary).In the lexicon for the most part we have to introduce the positive, negative nonpartisan and things. Every single huge datum and information mining ventures in view of the prepared information, without prepared information (introduction of words).So instatement of the prepared information is vital. In the self-learning framework, we are doing word institutionalization, here we are not considering past, present and future status of the words, just we are thinking about the word.

- **Sentiment Analysis Module**

In this module, preprocessed tweets are brought from the database one by one. In the first place we require check one by one watchword whether that catchphrase is thing are not, if thing we will expel it from the specific tweet. After that the rest of the watchwords checked with assessment compose, regardless of whether that catchphrases are certain opinion or negative conclusion or impartial feeling. The rest of the watchwords in the tweet which does not has a place with any of the supposition will be relegated transitory conclusion in light of the more check of positive, negative and impartial. In the second cycle if the reaming word crosses the limit of positive, negative or nonpartisan, that watchword forever included as development in the lexicon. At long last opinion of the tweet is recognized in light of the positive, negative and impartial words in the specific tweet.

4.7 System Design

4.7.1 The activities of the Design process:

1. Interface outline portrays the structure and association of the UI. Incorporates a portrayal of screen format, a meaning of the methods of collaboration, and a depiction of route components. Interface Control instruments to actualize route choices, the originator chooses frame one of various connection system;

a. Navigation menus

b. Graphic symbols

c. Graphic pictures

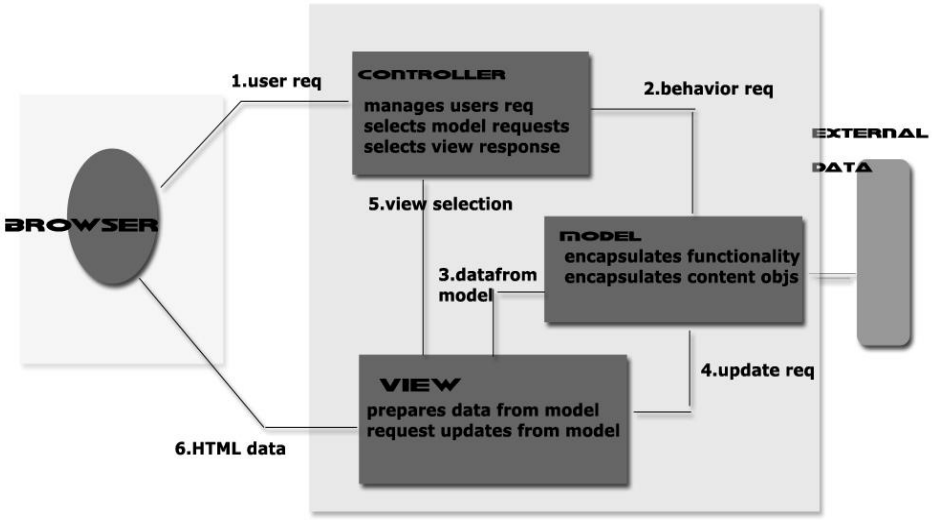
Interface Design work process the work process starts with the ID of client, undertaking, and natural necessities. When client errands have been distinguished, client situations are made and broke down to characterize an arrangement of interface protests and activities.

2. Content plan characterizes the design, structure, and blueprint for all substance that is exhibited as a component of the Web App. Builds up the connections between items.

3. Navigation outline speaks to the navigational stream between substance objects and for all Web App capacities.

4. Architecture outline distinguishes the general hypermedia structure for the WebApp. Engineering configuration is attached to the objectives set for a WebApp, the substance to be exhibited, the clients will visit, and the route logic that has been made.

- Content engineering, centers around the way in which content protests and organized for introduction and route.
- Web App design, addresses the way in which the application is structure to oversee client communication, handle inner preparing errands, impact route, and present substance. Web App design is characterized inside the setting of the advancement condition in which the application is to be actualized.



J2EE uses MVC Architecture

4.8 System Architecture

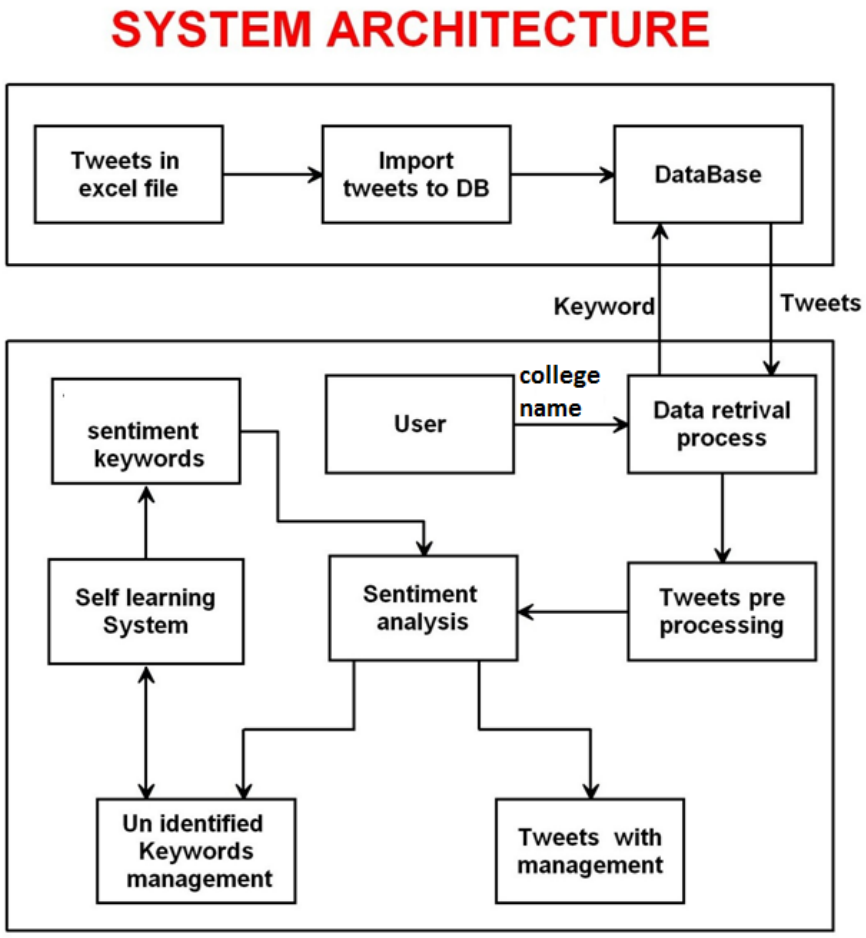


Figure 4.8.1 System Architecture

4.9 Use Case Diagram

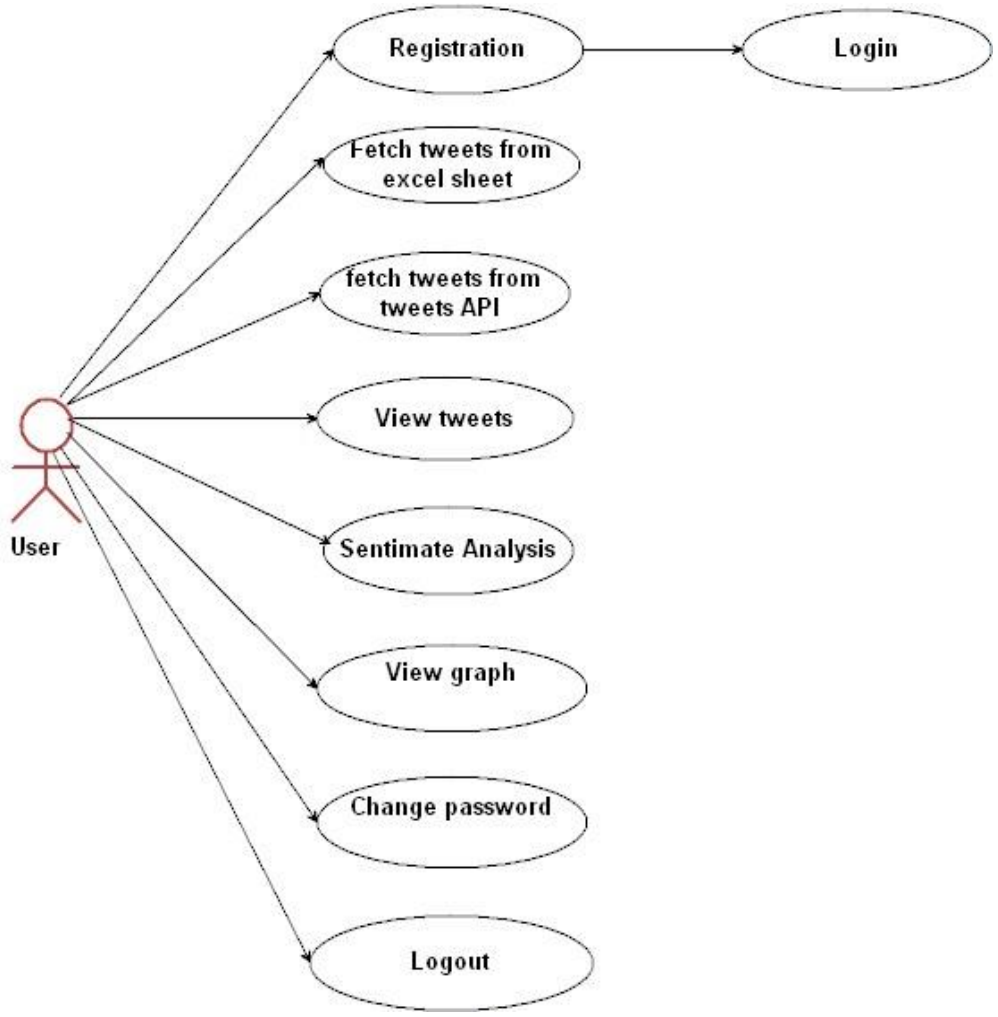


Figure 4.9.1 Use Case Diagram

4.10 Data Flow Diagram

4.10.1DFD User Session

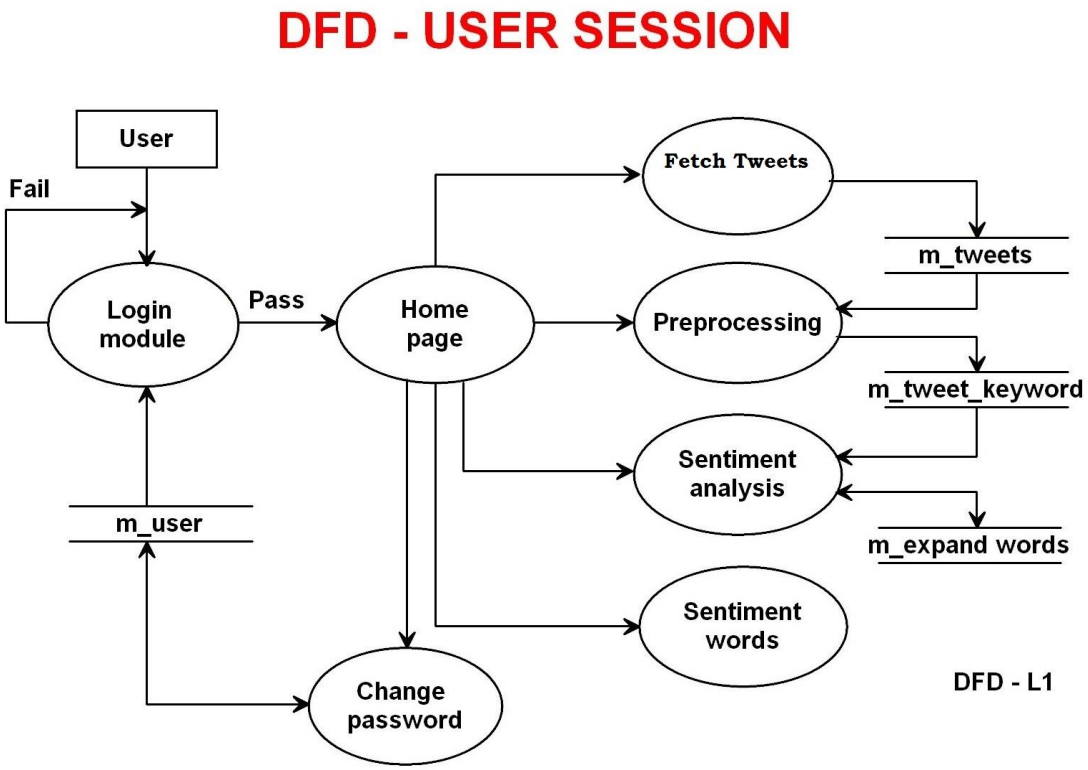


Figure 4.10.1.1 DFD User Session Diagram

4.10.2 DFD Preprocessing

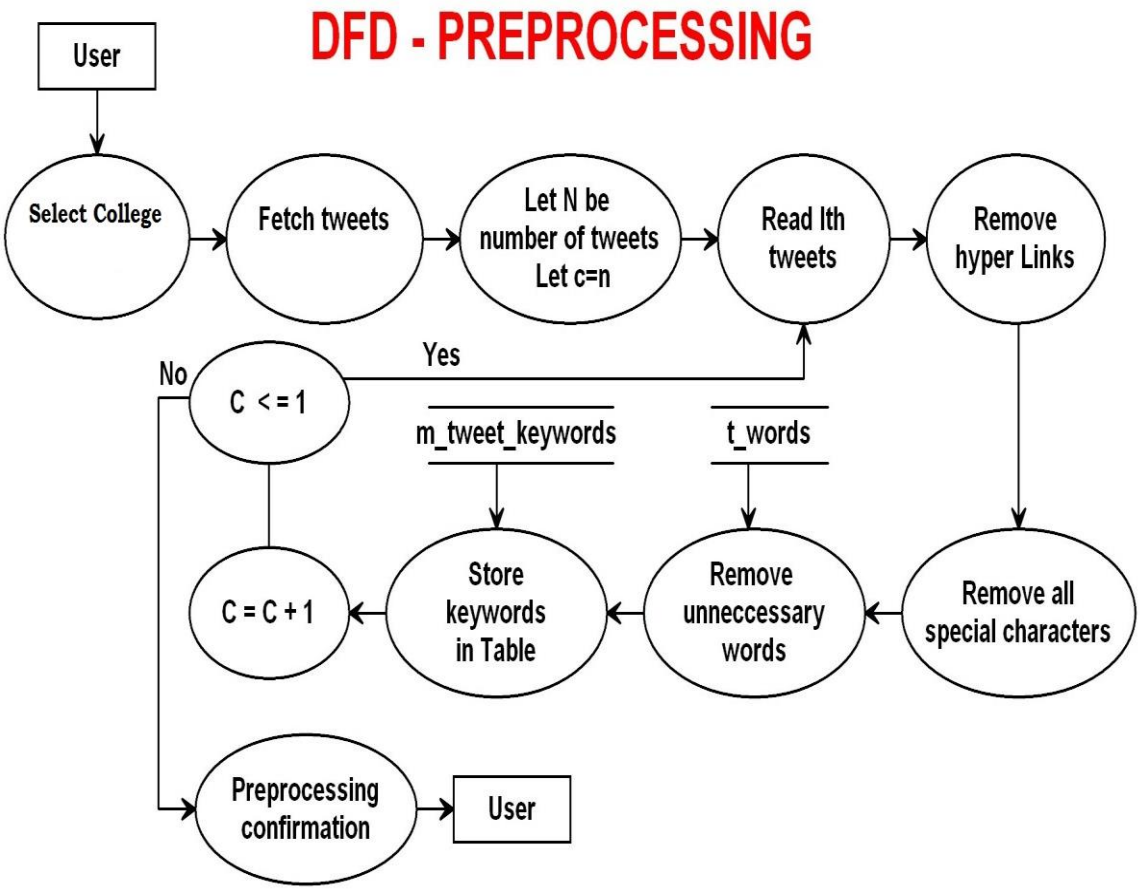


Figure 4.10.2.1 DFD Preprocessing

4.11 Sequence Diagram

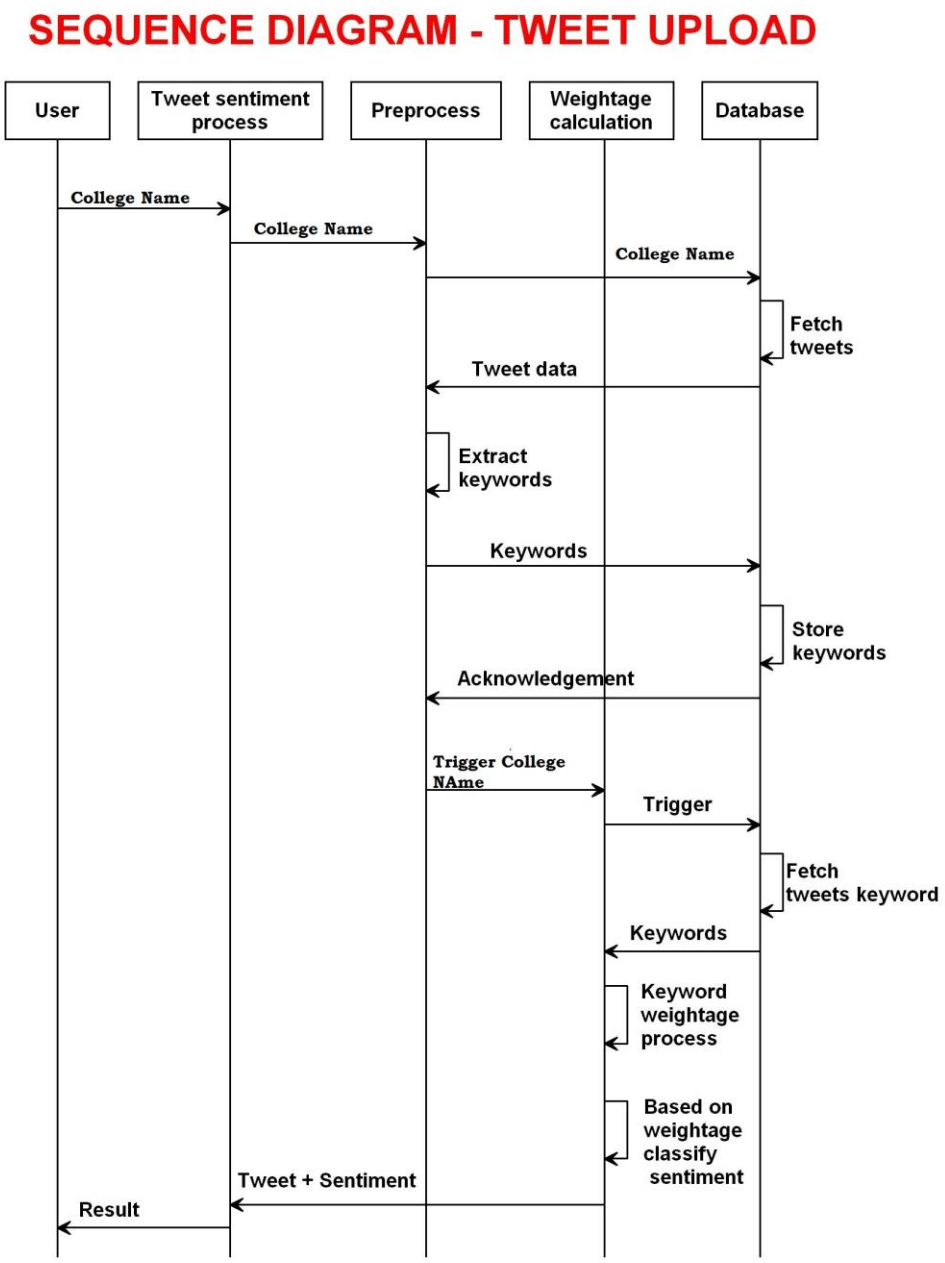


Figure 4.11.1 Sequence Diagram Upload Process

4.12 Context Analysis

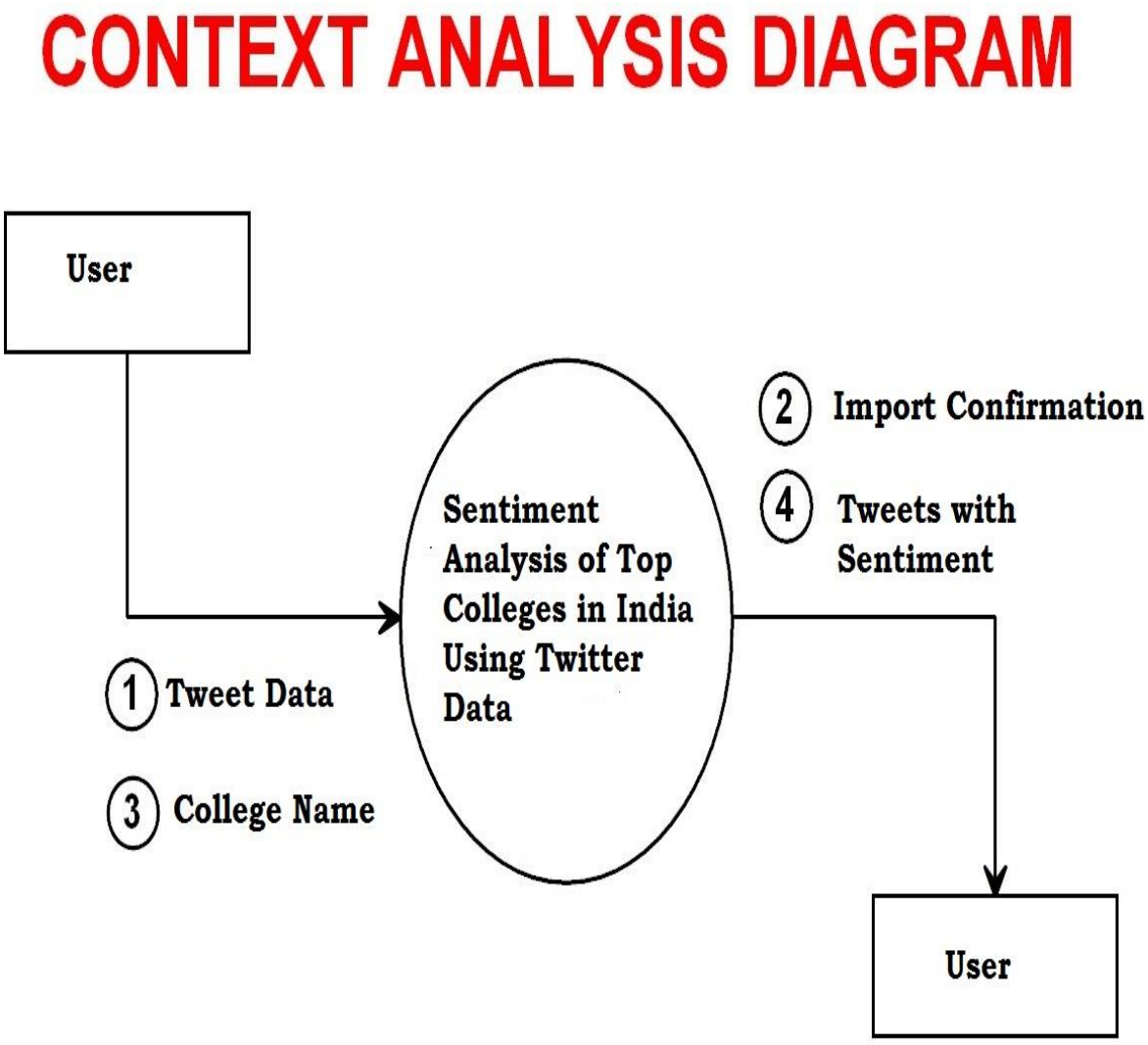


Figure 4.12.1 Context Analysis Diagram

4.13 Flow Chart Draw the flow chart with proper arrow pointing correct flow

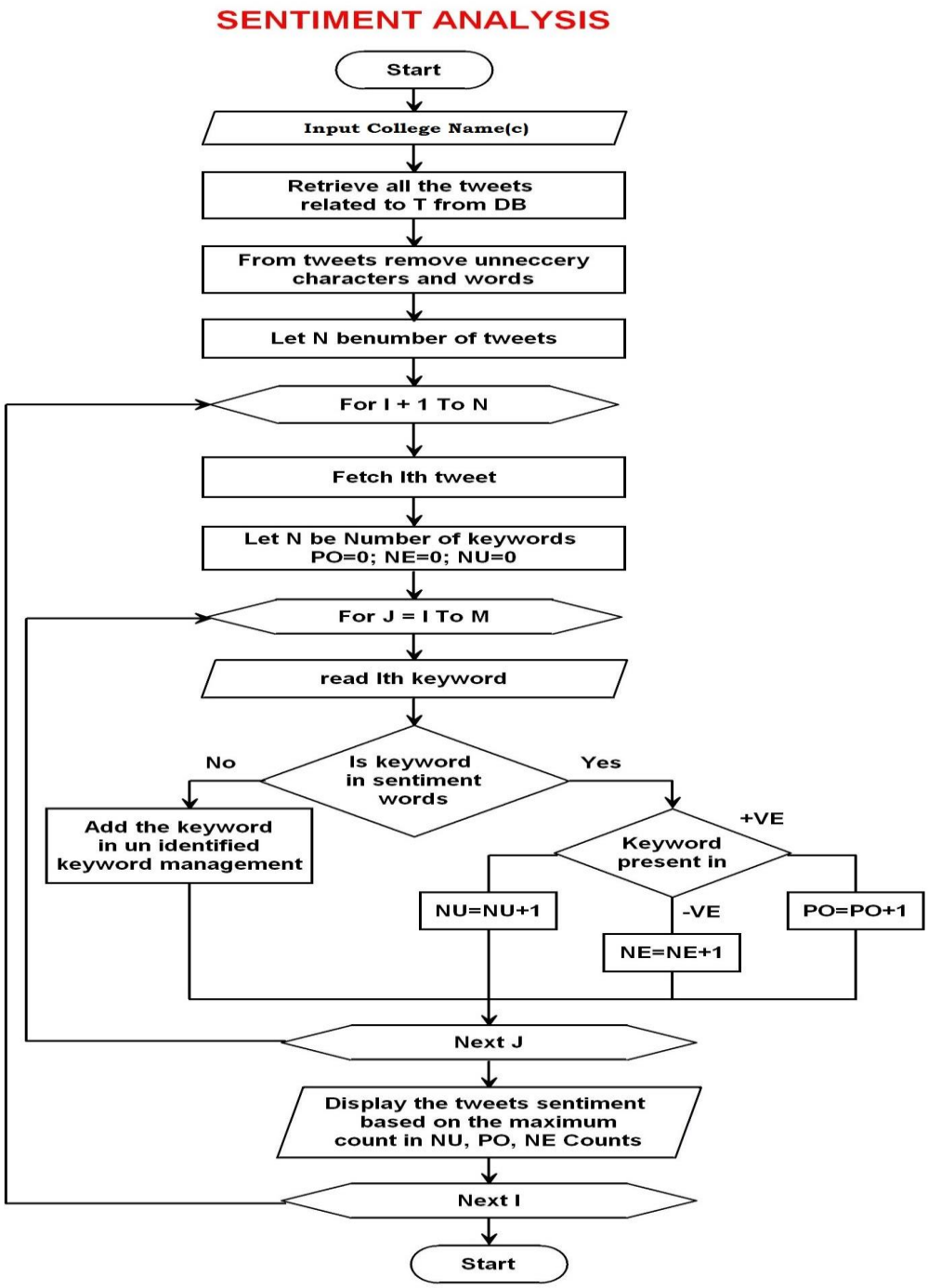


Fig.4.13.1 Flow Chart for Sentiment mining

CHAPTER 5

TESTING

5.1 Definition

Unit testing is an improvement system where software engineers make tests as they create programming. The tests are basic short tests that test practically of a specific unit or module of their code, for example, a class or capacity.

Utilizing open source libraries like unit, opponent and religious recluse it (for C, C++ and Java) these tests can be consequently run and any issues discovered rapidly. As the tests are created in parallel with the source unit test exhibits its rightness.

5.2 Validation and System Testing

Approval testing is a worry which covers with reconciliation testing. Guaranteeing that the application satisfies its detail is a noteworthy standard for the development of a coordination test. Approval testing additionally covers to a huge degree with System Testing, where the application is tried concerning its normal workplace. Thusly for some procedures no reasonable division amongst approval and framework testing can be made. Particular tests which can be performed in either or the two phases incorporate the accompanying.

- **Regression Testing:** Where this version of the software is tested with the automated test harness used with previous versions to ensure that the required features of the previous version are skill working in the new version.

- **Recovery Testing:** Where the software is deliberately interrupted in a number of ways off, to ensure that the appropriate techniques for restoring any lost data will function.

- **Security Testing:** Where unauthorized attempts to operate the software, or parts of it, attempted it might also include attempts to obtain access the data, or harm the software installation or even the system software. As with all types of security

determined will be able to obtain unauthorized access and the best that can be achieved is to make this process as difficult as possible.

- **Stress Testing:** Where abnormal demands are made upon the software by increasing the rate at which it is asked to accept, or the rate t which it is asked to produce information. More complex tests may attempt to crate very large data sets or cause the software to make excessive demands on the operating system.

- **Performance testing:** Where the performance requirements, if any, are checked. These may include the size of the software when installed, type amount of main memory and/or secondary storage it requires and the demands made of the operating when running with normal limits or the response time.

- **Usability Testing:** The process of usability measurement was introduced in the previous chapter. Even if usability prototypes have been tested whilst the application was constructed, a validation test of the finished product will always be required.

- **Alpha and beta testing:** This is where the software is released to the actual end users. An initial release, the alpha release, might be made to selected users who be expected to report bugs and other detailed observations back to the production team. Once the application changes necessitated by the alpha phase can be made to larger more representative set users, before the final release is made to all users.

  The final process should be a **Software audit** where the complete software project is checked to ensure that it meets production management requirements. This ensures that all required documentation has been produced, is in the correct format and is of acceptable quality. The purpose of this review is: firstly to assure the quality of the production process and by implication construction phase commences. A formal hand over from the development team at the end of the audit will mark the transition between the two phases.

5.3 Integration Testing

Integration Testing can continue in various diverse ways, which can be comprehensively portrayed as best down or base up. In top down reconciliation testing the abnormal state control schedules are tried to begin with, conceivably with the center level control structures exhibit just as stubs. Subprogram stubs were introduced in section2 as fragmented subprograms which are just present to permit the higher. Level control schedules to be tried.

Top down testing can continue in a profundity first or an expansiveness first way. For profundity first reconciliation every module is tried in expanding subtle element, supplanting an ever increasing number of levels of detail with real code as opposed to stubs. On the other hand expansiveness initially would handle by refining every one of the modules at a similar level of control all through the application .by and by a blend of the two methods would be utilized. At the underlying stages every one of the modules may be just halfway useful, potentially being actualized just to manage non-mistaken information. These eventual tried in expansiveness first way, yet over some stretch of time each future supplanted with progressive refinements which were nearer to the full usefulness. This permits profundity initially testing of a module to be performed at the same time with expansiveness initially testing of the considerable number of modules.

The other significant classification of mix testing is Bottom Up Integration Testing where an individual module is tried frame a test outfit. Once an arrangement of individual module have been tried they are then consolidated into an accumulation of modules ,known as assembles, which are then tried by a moment test saddle. This procedure can proceed until the point that the fabricate comprises of the whole application. By and by a blend of best down and base up testing would be utilized. In a substantial programming venture being produced by various sub-groups, or a littler venture where diverse modules were worked by people. The sub groups or people would lead base up testing of the modules which they were developing before discharging them to a coordination group which would gather them together for top-down testing.

5.4 Unit Testing

Unit testing manages testing a unit overall. This would test the association of many capacities yet bind the test to one unit. The correct extent of a unit is left to understanding. Supporting test code, once in a while called Scaffolding, might be important to help an individual test. This sort of testing is driven by the engineering and usage groups. This concentration is additionally called discovery testing on the grounds that exclusive the points of interest of the interface are unmistakable to the test. Limits that are worldwide to a unit are tried here.

In the development business, platform is a transitory, simple to amass and dismantle, outline set around a working to encourage the development of the building. The development specialists initially construct the platform and afterward the building. Later the framework is expelled, uncovering the finished building likewise, in programming testing, one specific test may require some supporting programming. This product builds up can a right assessment of the test occur. The platform programming may build up state and qualities for information structures and giving sham outside capacities to the test. Distinctive platform programming might be required frame one test to another test. Platform programming seldom is considered piece of the framework.

A few times the platform programming winds up noticeably bigger than the framework programming being tried. Typically the platform programming is not of an indistinguishable quality from the framework programming and oftentimes is very delicate. A little change in test may prompt considerably bigger changes in the framework.

Inner and unit testing can be computerized with the assistance of scope apparatuses. Dissects the source code and created a test that will execute each option string of execution. Regularly, the scope device is utilized as a part of a marginally extraordinary way. To begin with the scope instrument is utilized to enlarge the source by putting data prints after each line of code. At that point the testing suite is executed creating a review trail. This review trail is broke down and reports the percent of the aggregate framework code executed amid the test suite. On the off chance that the scope is high and the untested source lines are of low effect to the framework's general quality, at that point not any more.

Conclusion

This segment deals with a couple sorts of testing, for instance, module testing that is a framework for testing the exact working of a particular unit of the source code. This is similarly suggested as unit testing. It also provides a brief knowledge regarding different sorts of blend testing in that solitary programming units are joined and attempted as a social event. This part similarly focuses on ensuring nature of the item.

CHAPTER 6

RESULT

The accompanying depictions layout the outcomes or yields that we are going to get once regulated execution of the considerable number of modules of the framework.

The general assessment got from the dataset with respect to the three universities AIIMS, IIT and NIT were, as per the following: a sum of 371 tweets were viewed as positive, 231 as negative and 542 as impartial for AIIMS. 612 tweets were named positive, 332 as negative and 798 as unbiased for IIT. 542 tweets were considered as positive, 327 as negative and 924 as unbiased for NIT. As recently referenced, a greatness of 0 was considered as impartial valence, more prominent than 0 was considered as positive valence while less than zero was considered as negative valence. Table I shows insights acquired.

TABLE I. Statistics on the Sentiments Extracted From Tweets in Existing System

College	**Ratio of positive to negative tweets**	**Average positive sentiment**
NIT	1.43	2.94
IIT	1.55	2.93
AIIMS	1.73	4.56

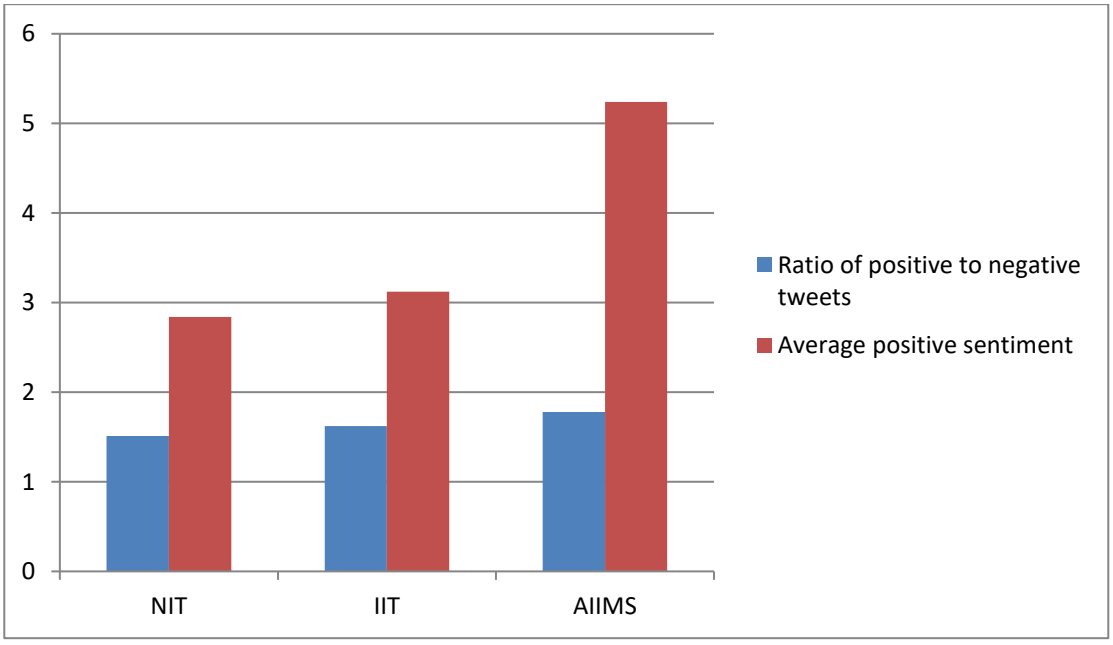


Figure 2. Shows the Sentiments Extracted From Tweets in Existing System.

TABLE II. Statistics on the Sentiments Extracted From Tweets in Proposed System

College	**Ratio of positive to negative tweets**	**Average positive sentiment**
NIT	1.51	2.84
IIT	1.62	3.12
AIIMS	1.78	5.24

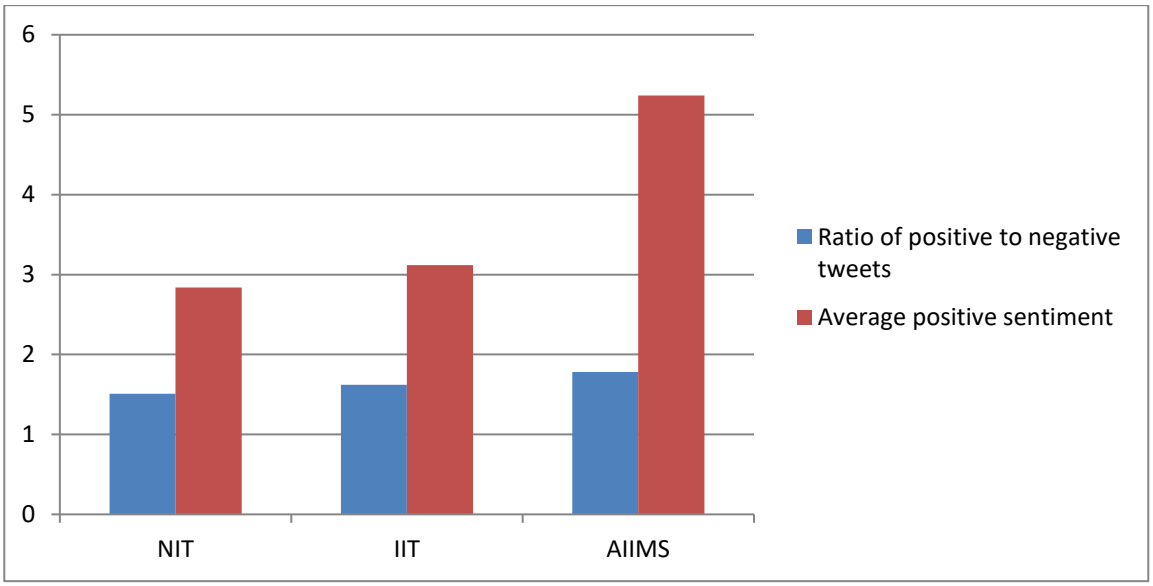


Figure 3. Shows the Sentiments Extracted From Tweets in Proposed System.

AIIMS had the most noteworthy positive normal assumption and the proportion for positive to negative tweets. This means the perception that the positive tweets about AIIMS are progressively positive in the greatness of their feeling and furthermore shows that AIIMS is discussed decidedly more than it is discussed adversely the most among the three organizations. The expectations made by the AI calculations demonstrated high exactness. For estimating exactness, ROC bends were built which plot the genuine positive rate as an element of the bogus positive rate at different edge settings. Basically, genuine positive rate portrays the quantity of tests anticipated to be sure which were likewise positive in fact. It is registered as the proportion of genuine positives to add up to positives. While, false positive rate means the quantity of tests which were really negative, yet were anticipated to be sure and is characterized as the proportion of false positives to add up to negatives.

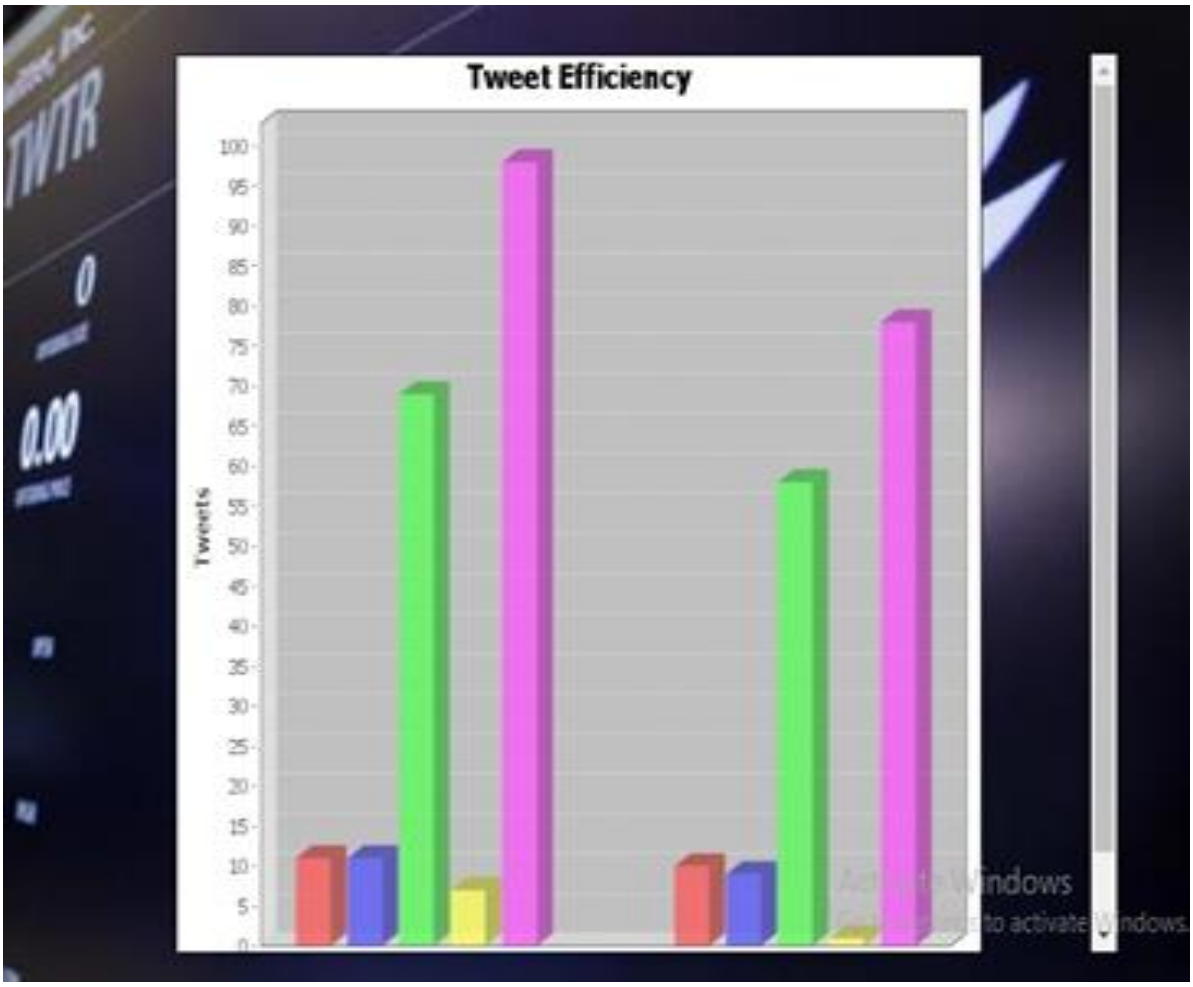
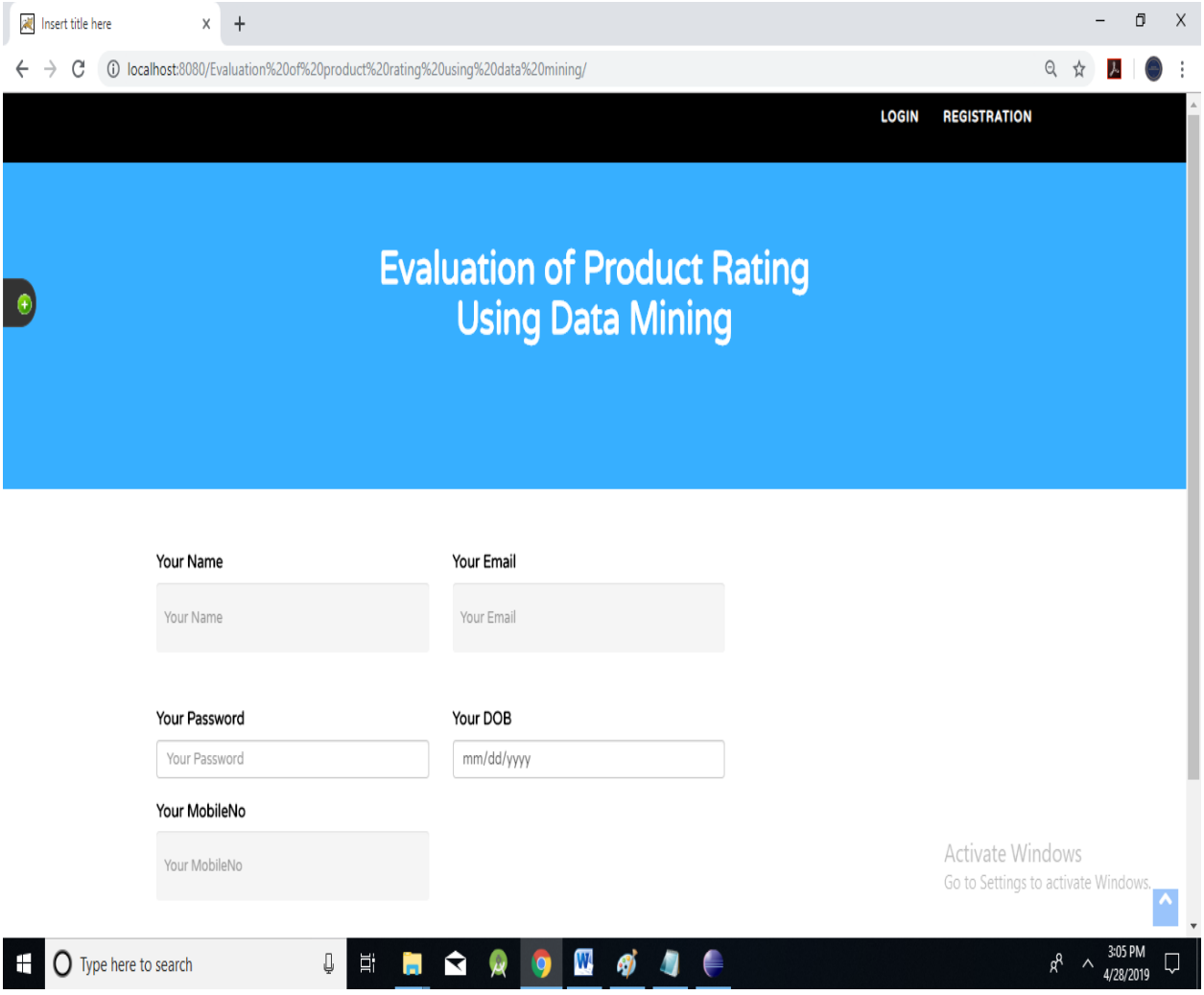


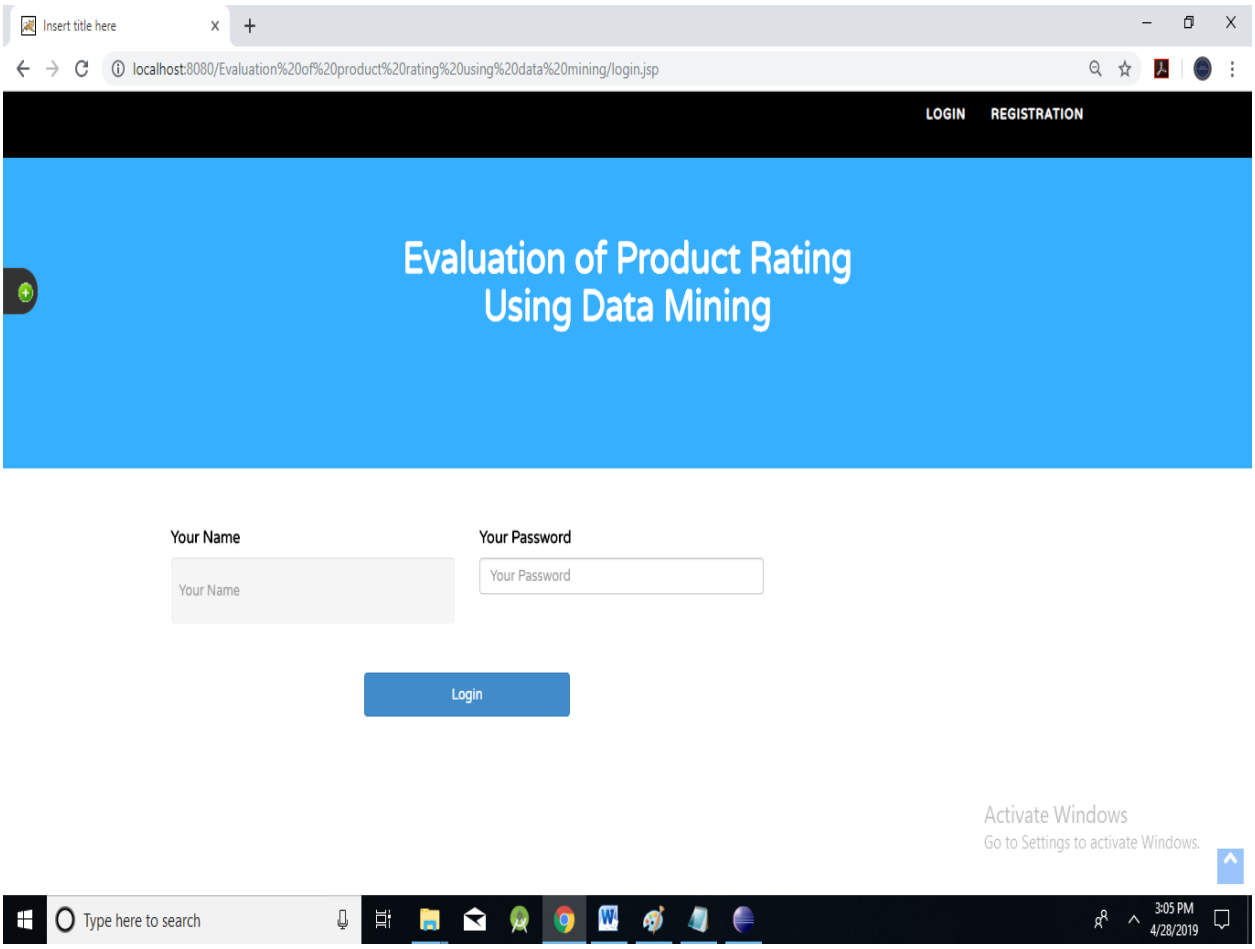
Figure 4.Shows the tweets efficiency graph.

Screenshot Results



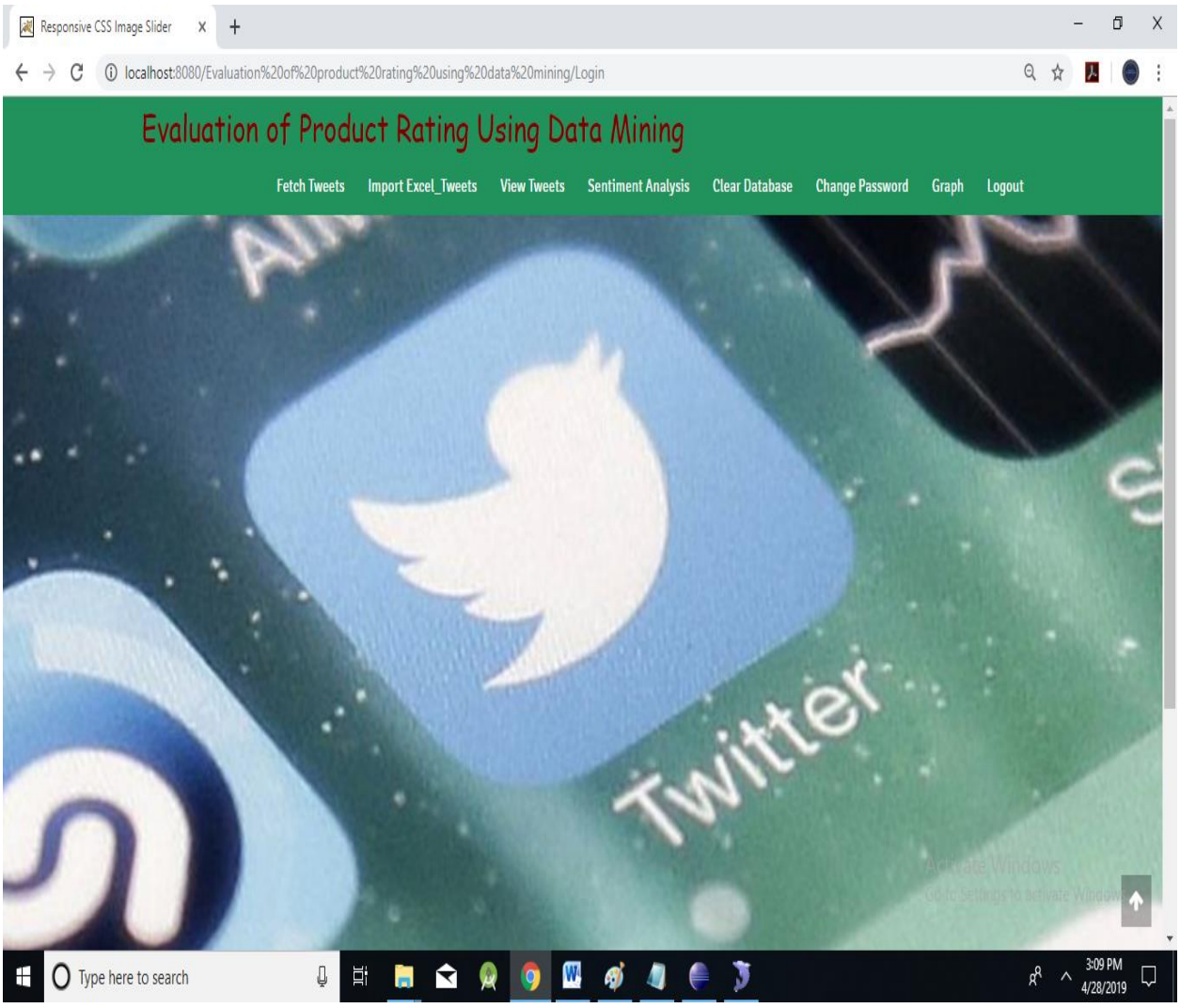
Screenshot-1 Registration Page

The screenshot-1 shows the Registration page.



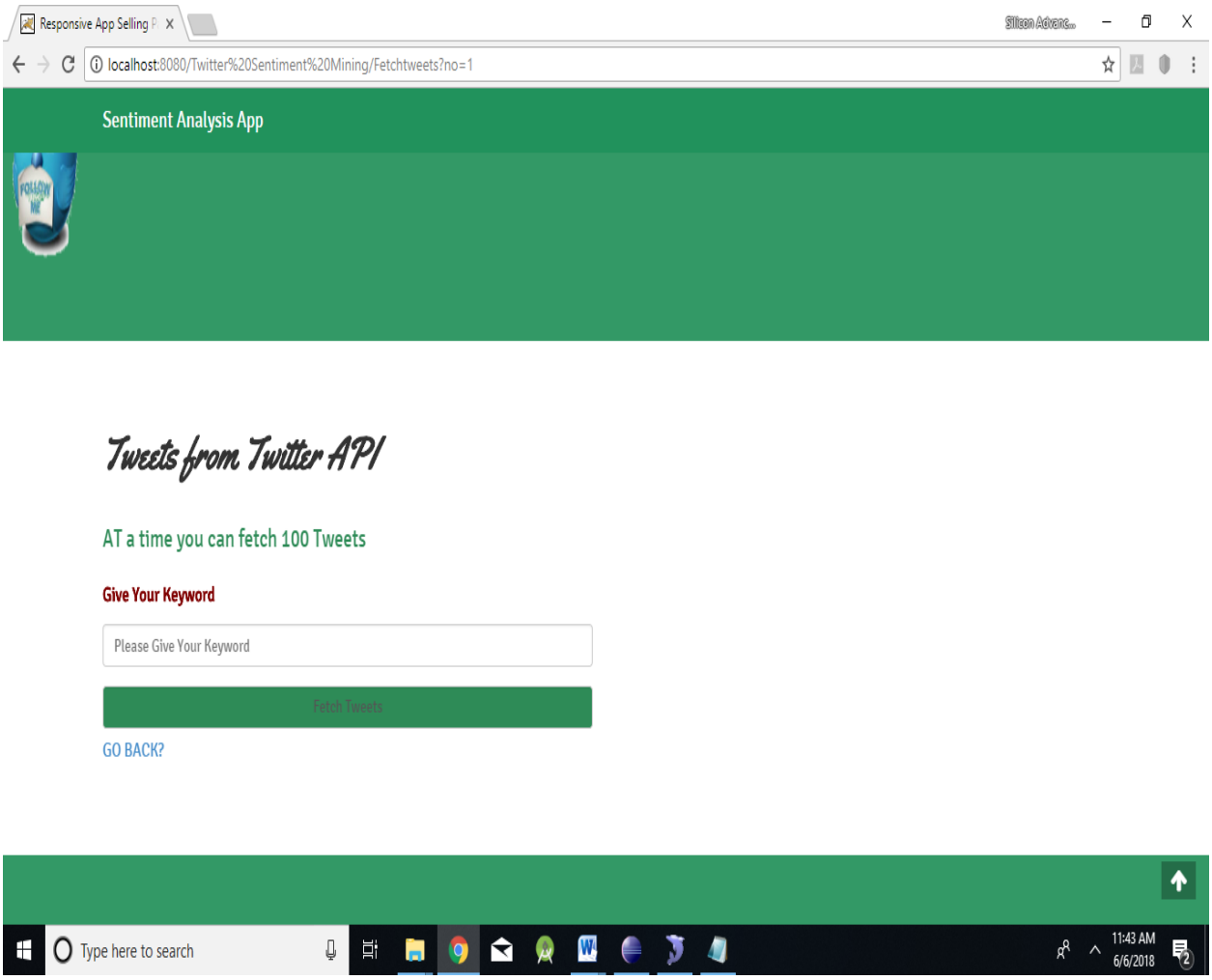
Screenshot-2 Login Page

The screenshot-2 shows the login page.



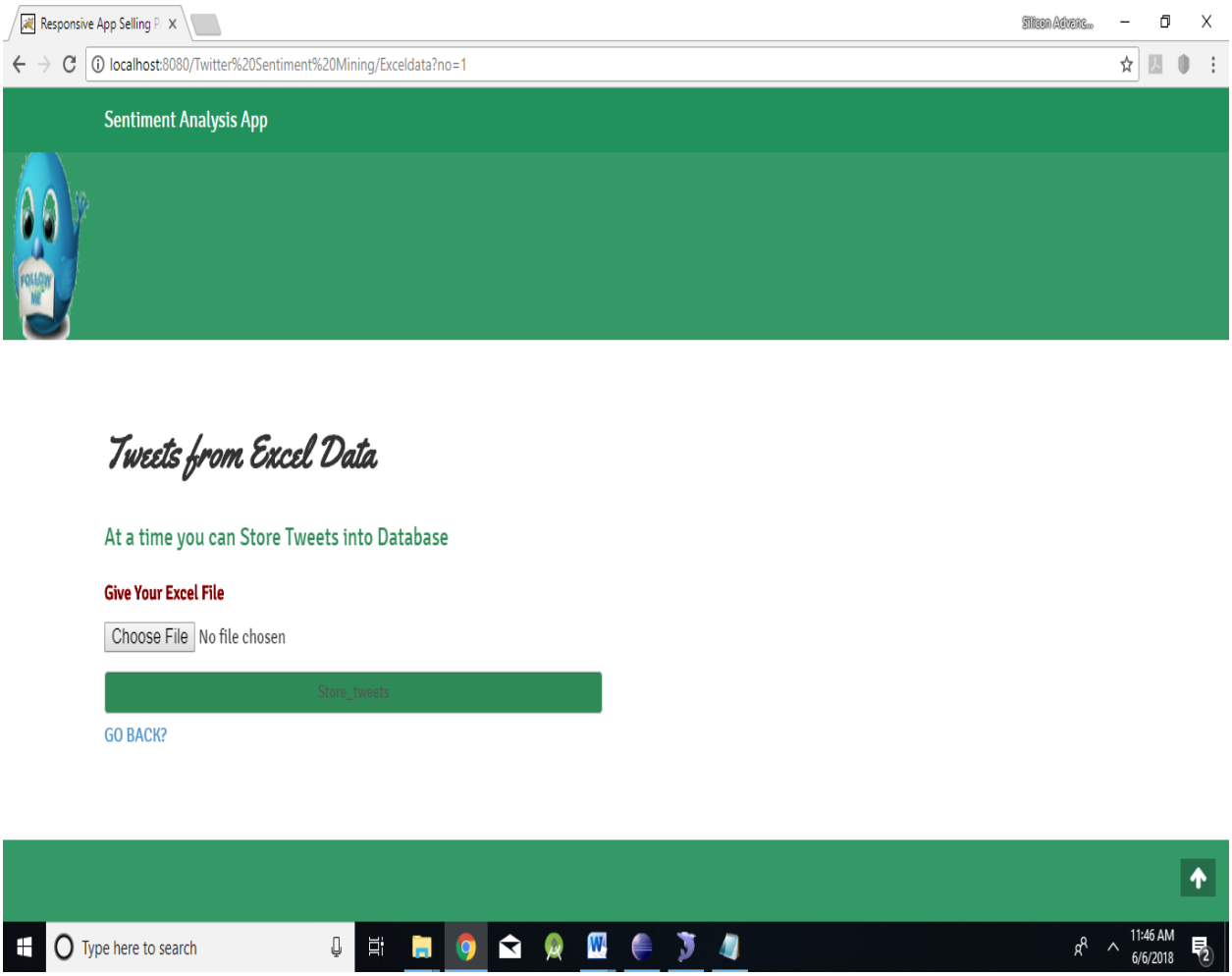
Screenshot-3 Home Page

The screenshot-3 shows the home page details.



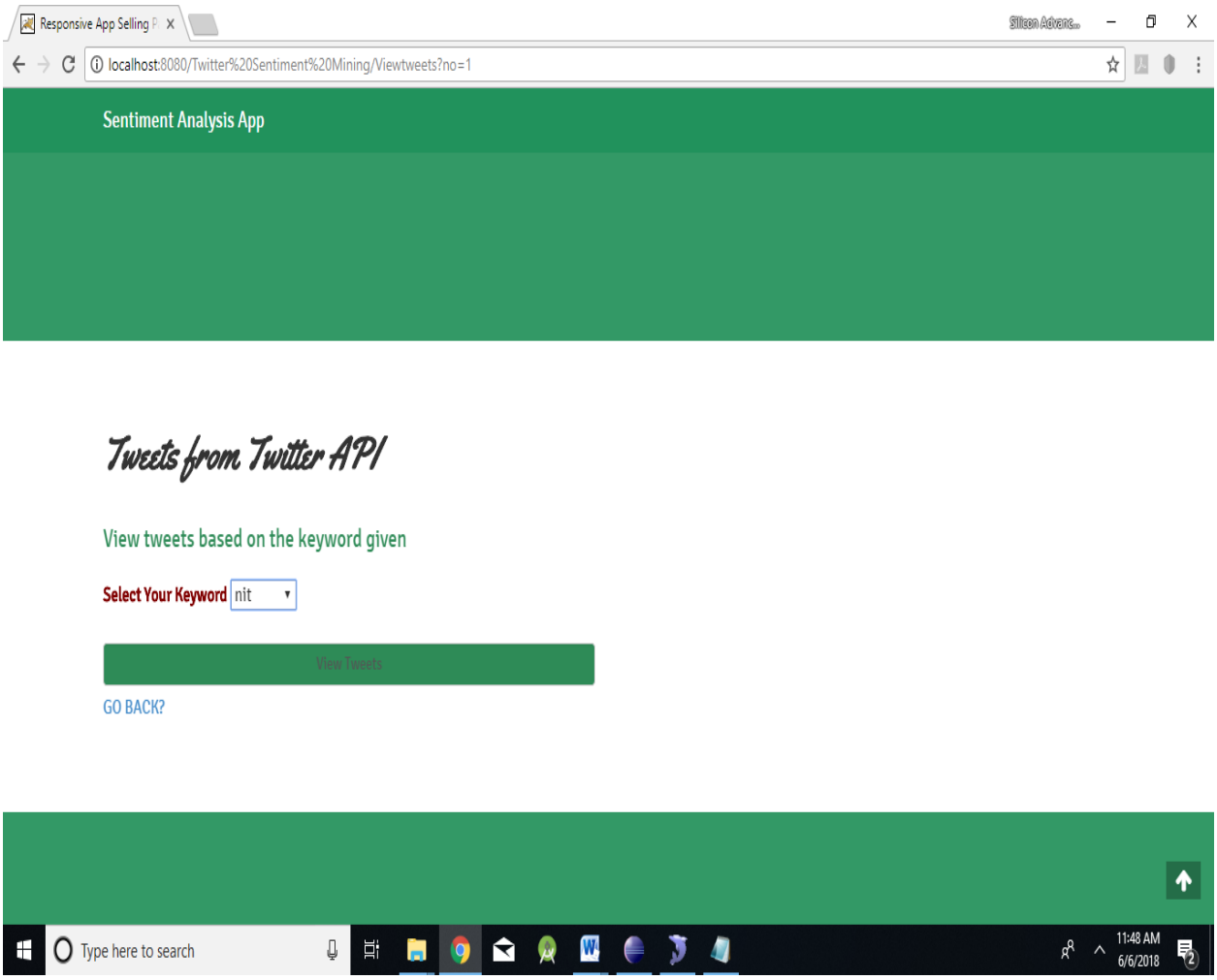
Screenshot-4 Tweets from twitter API

The screenshot-4 shows the Tweets from twitter API



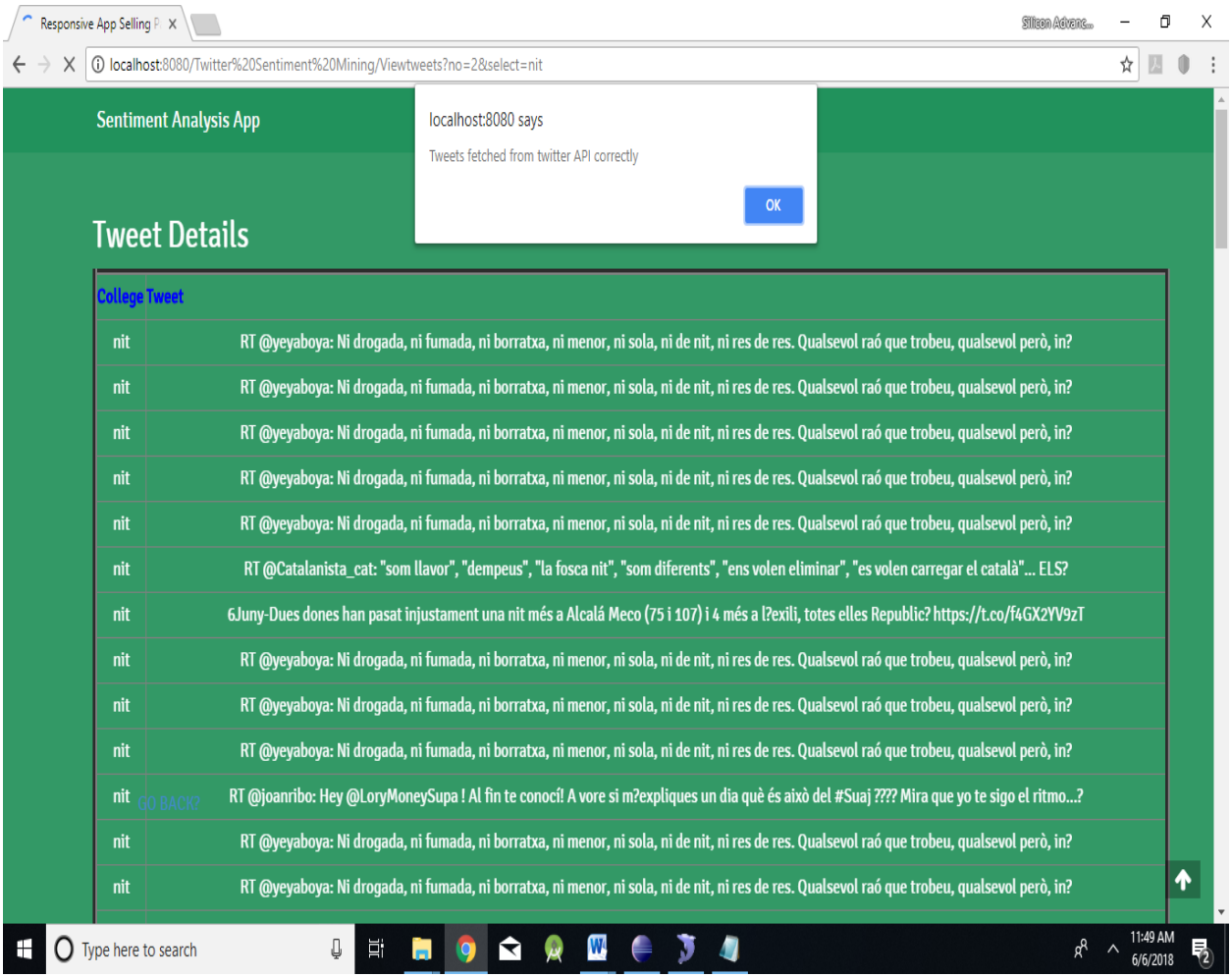
Screenshot-5 Tweets from excel data

The screenshot-5 shows the tweets form excel data.



Screenshot-6 View Tweets

The screenshot-6 shows the tweet details



localhost:8080 says

Tweets fetched from twitter API correctly

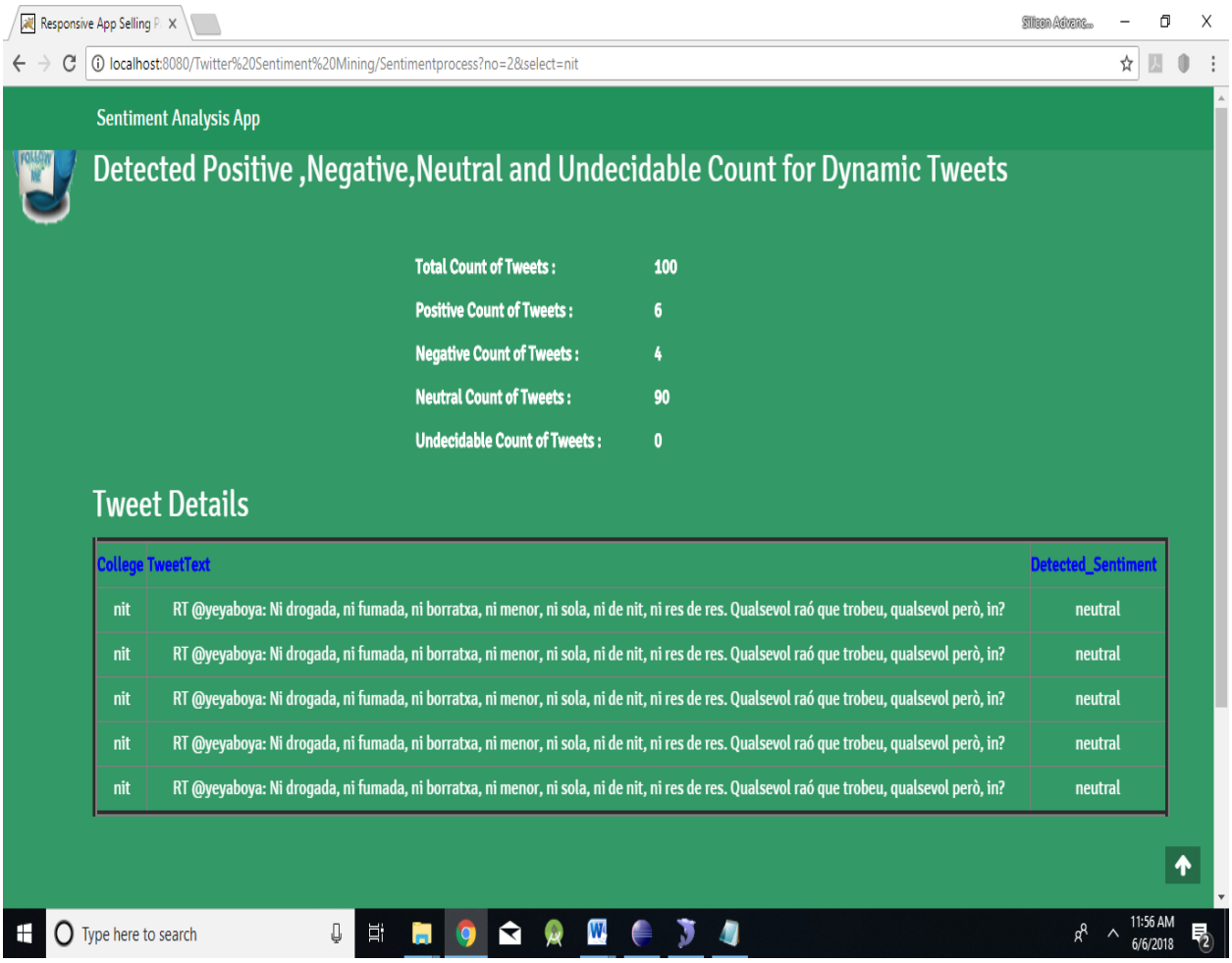
Tweet Details

College Tweet

nit	RT @yeyaboya: Ni drogada, ni fumada, ni borratxa, ni menor, ni sola, ni de nit, ni res de res. Qualsevol raó que trobeu, qualsevol però, in?
nit	RT @yeyaboya: Ni drogada, ni fumada, ni borratxa, ni menor, ni sola, ni de nit, ni res de res. Qualsevol raó que trobeu, qualsevol però, in?
nit	RT @yeyaboya: Ni drogada, ni fumada, ni borratxa, ni menor, ni sola, ni de nit, ni res de res. Qualsevol raó que trobeu, qualsevol però, in?
nit	RT @yeyaboya: Ni drogada, ni fumada, ni borratxa, ni menor, ni sola, ni de nit, ni res de res. Qualsevol raó que trobeu, qualsevol però, in?
nit	RT @yeyaboya: Ni drogada, ni fumada, ni borratxa, ni menor, ni sola, ni de nit, ni res de res. Qualsevol raó que trobeu, qualsevol però, in?
nit	RT @Catalanista_cat: "som llavor", "dempeus", "la fosca nit", "som diferents", "ens volen eliminar", "es volen carregar el català"... ELS?
nit	6Juny-Dues dones han pasat injustament una nit més a Alcalá Meco (75 i 107) i 4 més a l?exili, totes elles Republic? https://t.co/f4GX2YV9zT
nit	RT @yeyaboya: Ni drogada, ni fumada, ni borratxa, ni menor, ni sola, ni de nit, ni res de res. Qualsevol raó que trobeu, qualsevol però, in?
nit	RT @yeyaboya: Ni drogada, ni fumada, ni borratxa, ni menor, ni sola, ni de nit, ni res de res. Qualsevol raó que trobeu, qualsevol però, in?
nit	RT @yeyaboya: Ni drogada, ni fumada, ni borratxa, ni menor, ni sola, ni de nit, ni res de res. Qualsevol raó que trobeu, qualsevol però, in?
nit	RT @joanribo: Hey @LoryMoneySupa ! Al fin te conocí! A vore si m?expliques un dia què és això del #Suaj ???? Mira que yo te sigo el ritmo...?
nit	RT @yeyaboya: Ni drogada, ni fumada, ni borratxa, ni menor, ni sola, ni de nit, ni res de res. Qualsevol raó que trobeu, qualsevol però, in?
nit	RT @yeyaboya: Ni drogada, ni fumada, ni borratxa, ni menor, ni sola, ni de nit, ni res de res. Qualsevol raó que trobeu, qualsevol però, in?

Screenshot-7 Tweets details

The screenshot-7shows the Tweets details.



Sentiment Analysis App

Detected Positive ,Negative,Neutral and Undecidable Count for Dynamic Tweets

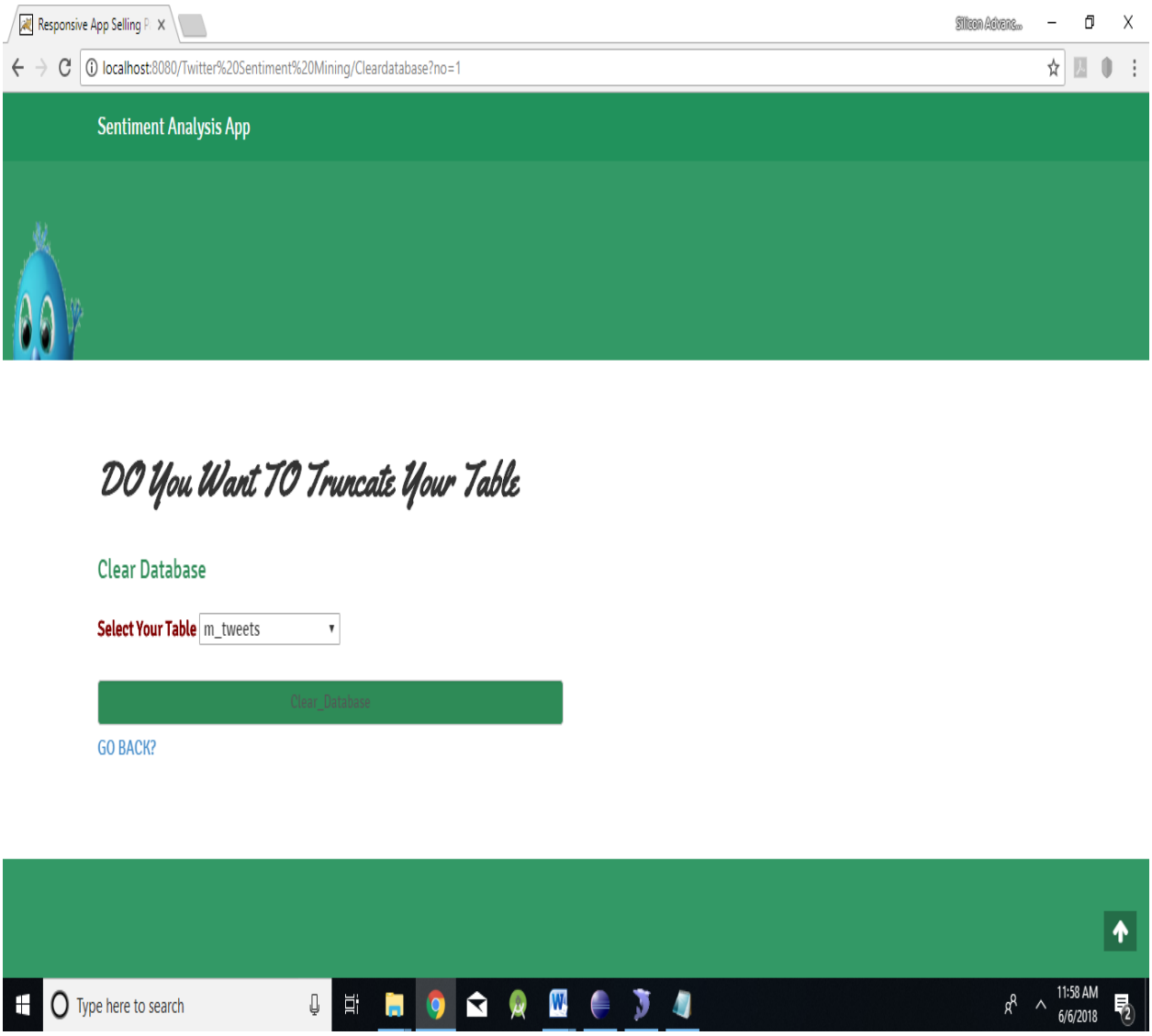
Total Count of Tweets :	100
Positive Count of Tweets :	6
Negative Count of Tweets :	4
Neutral Count of Tweets :	90
Undecidable Count of Tweets :	0

Tweet Details

College	TweetText	Detected_Sentiment
nit	RT @yeyaboya: Ni drogada, ni fumada, ni borratxa, ni menor, ni sola, ni de nit, ni res de res. Qualsevol raó que trobeu, qualsevol però, in?	neutral
nit	RT @yeyaboya: Ni drogada, ni fumada, ni borratxa, ni menor, ni sola, ni de nit, ni res de res. Qualsevol raó que trobeu, qualsevol però, in?	neutral
nit	RT @yeyaboya: Ni drogada, ni fumada, ni borratxa, ni menor, ni sola, ni de nit, ni res de res. Qualsevol raó que trobeu, qualsevol però, in?	neutral
nit	RT @yeyaboya: Ni drogada, ni fumada, ni borratxa, ni menor, ni sola, ni de nit, ni res de res. Qualsevol raó que trobeu, qualsevol però, in?	neutral
nit	RT @yeyaboya: Ni drogada, ni fumada, ni borratxa, ni menor, ni sola, ni de nit, ni res de res. Qualsevol raó que trobeu, qualsevol però, in?	neutral

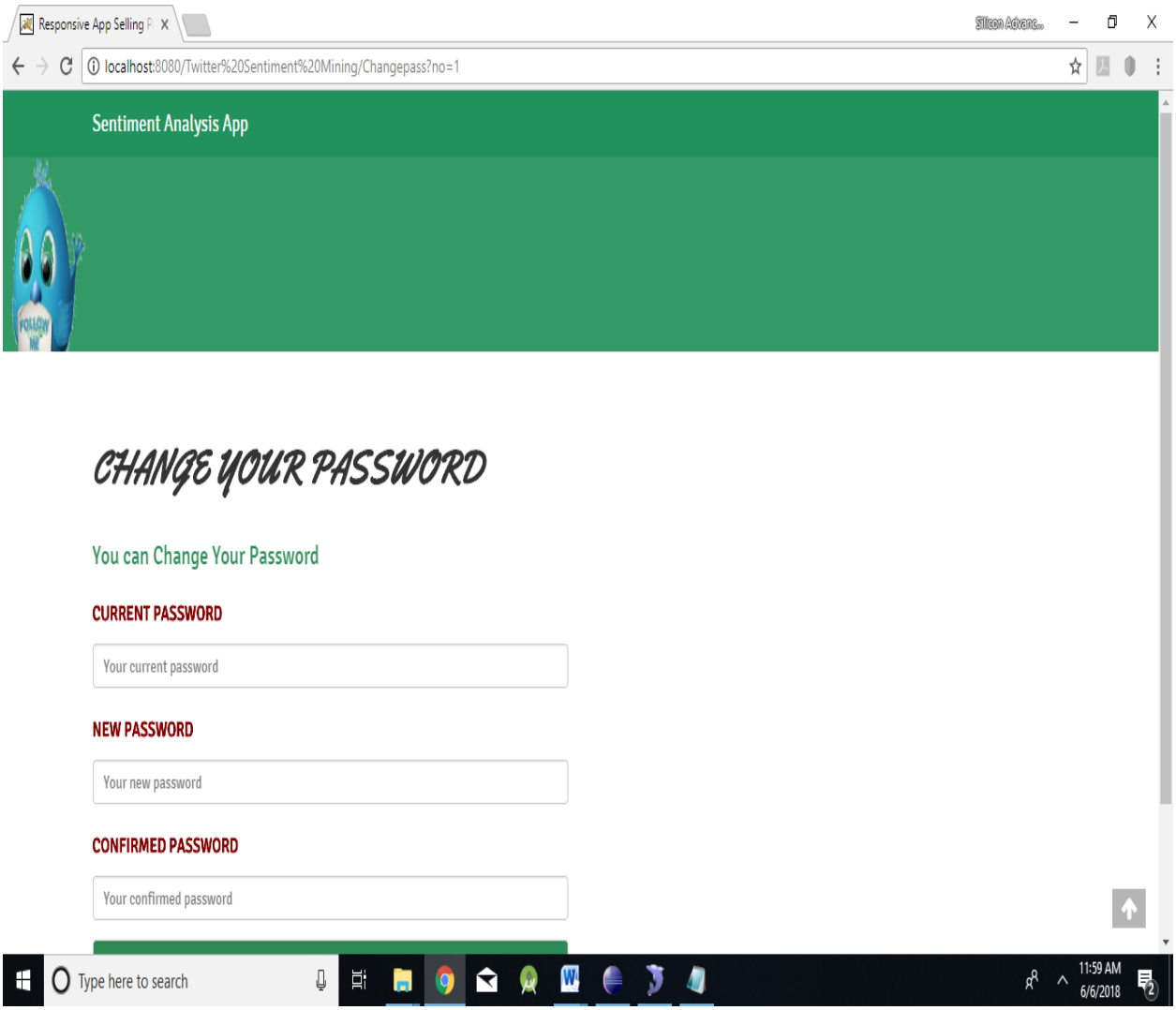
Screenshot-8 Sentiment Mining

The screenshot-8 shows the sentiment mining details Page.



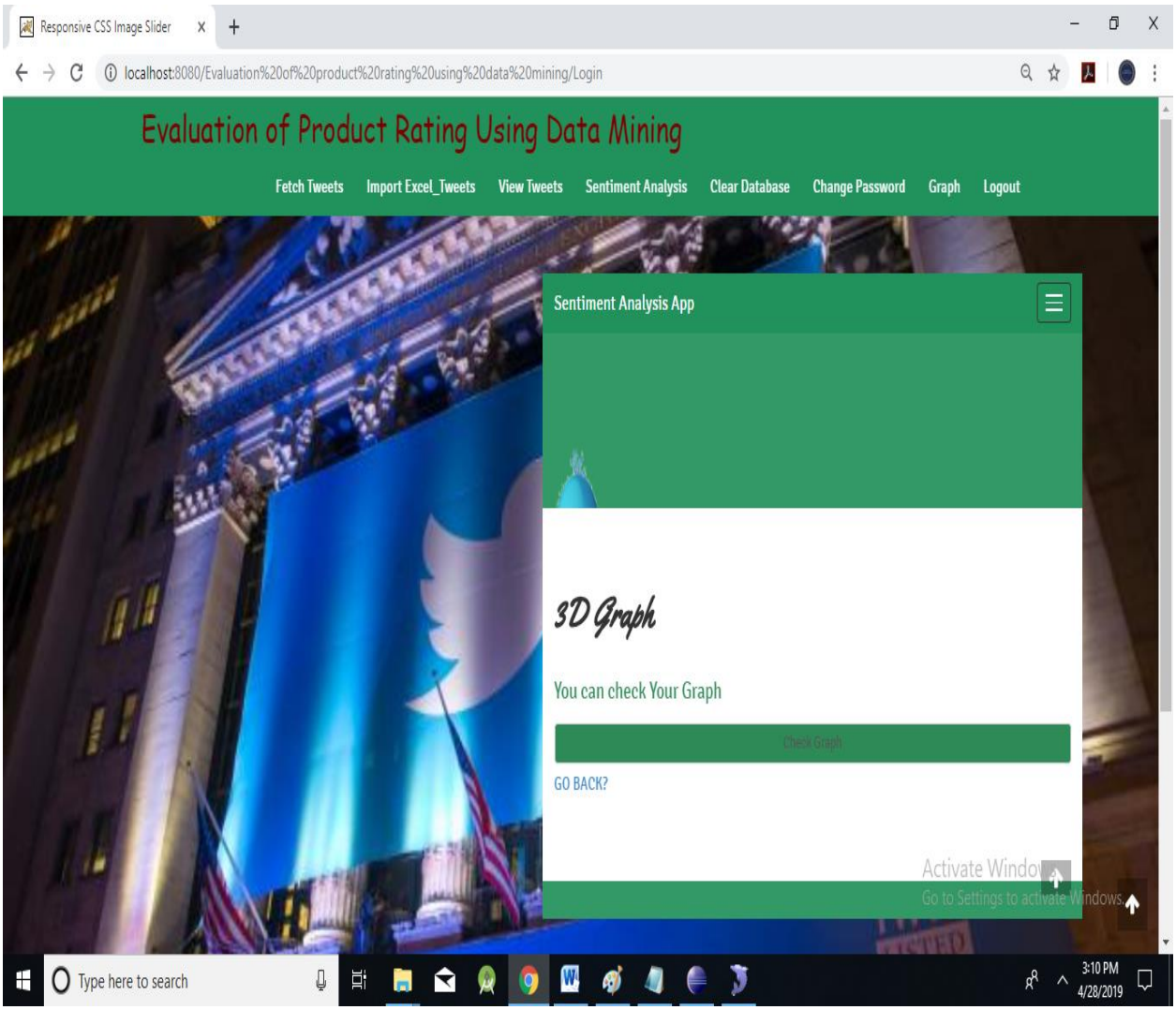
Screenshot-9 clear database Page

The screenshot-9 shows the Clear database page.



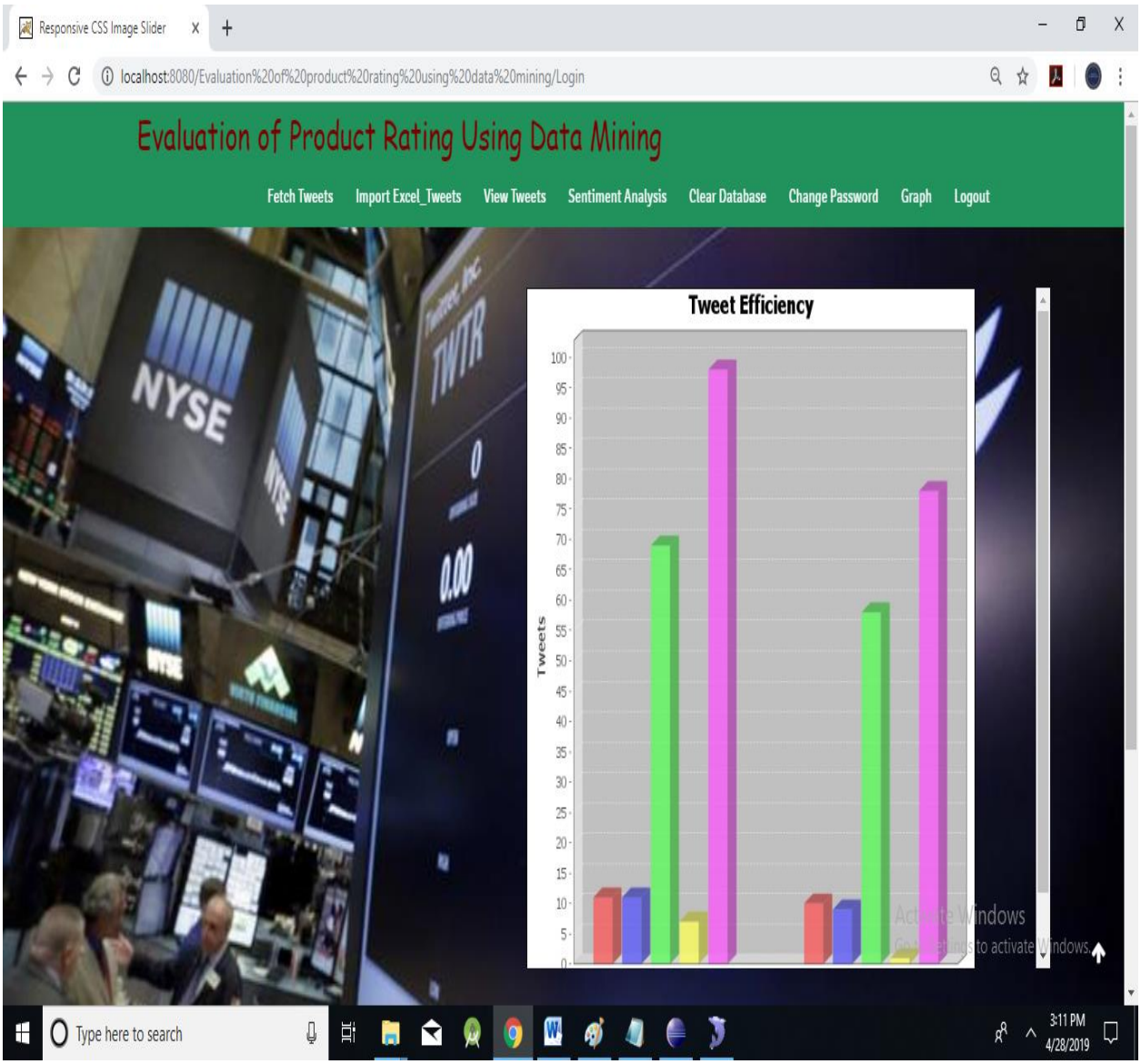
Screenshot-10 change password

The screenshot-10 shows the change password page.



Screenshot-11 Check graph page

The screenshot-11 shows the check graph page.



Screenshot-12 Tweet efficiency graph

The screenshot-12 shows the tweets efficiency graph.

CONCLUSION

Sentiment analysisis a compelling method for arranging the assessments figured by individuals with respect to any point, administration or item. Mechanization of this errand makes it less demanding to manage the monstrous measure of information being delivered by social sites like Twitter on an ongoing premise. Contrasted with the current framework, in view of the edge idea, cycle idea, watchwords are extended .Based on the extended words effectiveness of the framework is expanded.

REFERENCES

[1] Twitter Usage/Company Facts, https://about.twitter.com/company

[2] Posting a tweet, https://support.twitter.com/articles/15367-posting-atweet

[3] King R. A., Racherla P. and Bush V. D., What We Know and Don't Know about Online Word-of-Mouth: A Review and Synthesis of the Literature, Journal of Interactive Marketing, vol. 28, issue 3, pp. 167- 183, August 2014

[4] Ministry of Human Resource Development, http://mhrd.gov.in/statist

[5] India's Best Colleges, 2015, http://indiatoday.intoday.in/bestcolleges/2015/

[6] Arora D., Li K.F. and Neville S.W., Consumers' sentiment analysis of popular phone brands and operating system preference using Twitter data: A feasibility study, 29th IEEE International Conference on Advanced Information Networking and Applications, pp. 680-686, Gwangju, South Korea, March 2015

[7] Choi C., Lee J., Park G., Na J. and Cho W., Voice of customer analysis for internet shopping malls, International Journal of Smart Home: IJSH, vol. 7, no. 5, pp. 291-304, September 2013

[8] Kanakaraj M., Guddeti R M.R., Performance Analysis of Ensemble Methods on Twitter Sentiment Analysis using NLP Techniques, 9th IEEE International Conference on Semantic Computing, pp. 169-170, Anaheim, California, 2015

[9] Bahrainian S.-A., Dengel A., Sentiment Analysis and Summarization of Twitter Data", 16th IEEE International Conference on Computational Science and Engineering, pp. 227-234, Sydney, Australia, December 2013

[10] Pak A. and Paroubek P., Twitter as a Corpus for Sentiment Analysis and Opinion Mining, 7th International Conference on Language Resources and Evaluation, pp. 1320-1326, Valletta, Malta, May 2010

[11] Shahheidari S., Dong H., Bin Daud M.N.R., Twitter sentiment mining: A multidomain analysis, 7th IEEE International Conference on Complex, Intelligent and Software Intensive Systems, pp.144-149, Taichung, Taiwan, July 2013

[12] Neethu M. S. and Rajasree R., Sentiment Analysis in Twitter using Machine Learning Techniques, 4th IEEE International Conference on Computing, Communications and Networking Technologies, pp. 1-5, Tiruchengode, India, 2013

[13] Bespalov D., Bai B., Qi Y., and Shokoufandeh A., Sentiment classification based on supervised latent n-gram analysis, 20th ACM international conference on Information and knowledge management, pp. 375-382, New York, USA, 2011

[14] Jotheeswaran J. and Koteeswaran S., Decision Tree Based Feature Selection and Multilayer Perceptron for Sentiment Analysis, Journal of Engineering and Applied Sciences, vol. 10, issue 14, pp. 5883-5894, January 2015

[15] Socher R., et al, Recursive Deep Models for Semantic Compositionality Over a Sentiment Treebank, Conference on Empirical Methods in Natural Language Processing, Seattle, Washington, October 2013.

[16] dos Santos C. N. and Gatti M., Deep Convolutional Neural Networks for Sentiment Analysis of Short Texts, 25th International Conference on Computational Linguistics, pp. 69-78, Dublin, Ireland, August 2014.

[17] Segaran T. and Hammerbacher J., Beautiful Data: The Stories behindElegant Data Solutions, Beijing: O'Reilly, 2009

[18] Nielsen F.A., Making sense of microposts, Finn Årup Nielsen blog,https://finnaarupnielsen.wordpress.com/tag/sentimentanalysis/

[19] Koto F. and Adriani M., A Comparative Study on Twitter SentimentAnalysis: Which Features are Good?, Natural Language Processing andInformation Systems, Lecture Notes in Computer Science vol. 9103, pp.453-457, June 2015

[20] Ng A.Y., Jordan M. I., On Discriminative vs. Generative classifiers: Acomparison of logistic regression and naive Bayes, Advances in NeuralInformation Processing Systems vol. 14, pp. 841-848, 2002

[21] Karatzoglou A., Meyer D., Hornik K., Support vector machines in R.,Journal of Statistical Software, vol. 15, issue: 9, April 2006

[22] Rajendran S., Kalpana B., A Comparative Study and Choice of anAppropriate Kernel for Support Vector Machines., International Journalof Soft Computing and Engineering (IJSCE), vol. 1, issue: 5, November2011

[23] Abakar K. A. A., Yu C., Performance of SVM based on PUK kernel incomparison to SVM based on RBF kernel in prediction of yarn tenacity,Indian Journal of Fibre and Textile Research, vol. 39, pp. 55-59, March2014

[24] Salazar D. A., Velez J. I., Salazar J. C., Comparison between SVM and Logistic Regression: Which One is Better to Discriminate?, ColombianJournal of Statistics, Special Issue Biostatistics, vol. 35, no. 2, pp. 223-237, June 2012

[25] Hsu C.-W., Chang C.-C., Lin C.-J., A practical guide to support vectorclassification, National Taiwan University, Taipei, April 2010

[26] Girma H., A Tutorial on Support Vector Machine, Center ofExperimental Mechanics, University of Ljubljana, 2009

[27] Multilayer perceptron,https://en.wikipedia.org/wiki/Multilayer_perceptron

[28] Manning C. and Raghavan P., Introduction to information retrieval, NewYork: Cambridge University Press, 2008.

Appendix-1

Sample Coding

```
packagecom.sentiment;
importjava.io.IOException;
importjava.io.PrintWriter;
importjava.sql.ResultSet;
importjava.util.ArrayList;
importjavax.servlet.RequestDispatcher;
importjavax.servlet.http.HttpServlet;
importjavax.servlet.http.HttpServletRequest;
importjavax.servlet.http.HttpServletResponse;
importjavax.servlet.http.HttpSession;

importcom.dataa.Sentimentprocessexicution;

importjava.util.HashMap;
importjava.util.Map;
public class Sentimentprocess extends HttpServlet
{
	public void doGet(HttpServletRequestrequest,HttpServletResponse response)throws IOException
	{	String college="";
		ArrayList<String> data=new ArrayList<String>();
		String sensitivetype="";
		boolean flag77=false;
		booleanff=false;
		int positivecount11=0;
		int negativecount11=0;
		int undecidablecount11=0;
		int neutralcount11=0;

		inttotalcount =0;
		booleanfinalcount=false;
		try {
		System.out.println("its came Sentimentprocess fetch tweets");
```

```
int no=Integer.parseInt(request.getParameter("no"));

System.out.println("action is >>>>>>>>>>>>>"+no);

if(no==1)
{


RequestDispatcherrd=request.getRequestDispatcher("/jsp/sentiment.jsp");

rd.forward(request, response);


}

if(no==2)
{
HttpSession session = request.getSession(false);

if ( session.getAttribute( "waitPage" ) == null )

 {
session.setAttribute( "waitPage", Boolean.TRUE );
PrintWriter out1 = response.getWriter();
out1.println( "<html><head>" );
out1.println( "<title>Please Wait...</title>" );
out1.println( "<meta http-equiv=\"Refresh\" content=\"0\">"
);
out1.println( "</head><body bgcolor='#2E8B57'>" );
out1.println( "<br><br><br>" );
out1.print( "<center><imgsrc='animated.gif' width='400'
height='400'></img><br><br>");
out1.println("<font color='#fedcba' size='5'>");
out1.println( "Please Do not press Back or Refresh
button.....<br> " );
out1.println("<font color='#fedcba' size='5'>");
out1.println( "Sentiment Detection is going on " );
out1.println("<font color='#fedcba' size='5'>");
out1.println( "Please wait....</h1></center");
out1.close();
```

```
}
else
{

session.removeAttribute( "waitPage" );
college=request.getParameter("select");
System.out.println("college is
>>>>>>>>>>>>>>>>>>>>>>>>"+college);



HashMap<Integer, ArrayList<String>>map_tclaim=new
HashMap<Integer, ArrayList<String>>();
map_tclaim=AdminDAO.getm_tweets(college);

map_tclaim.size();

System.out.println(">>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>size"+map_tclaim.s
ize());
System.out.println("entry set
is>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>"+map_tclaim.entrySet());

if(map_tclaim.size()>0)
{

for(Map.Entry m4:map_tclaim.entrySet())
{
// System.out.println("entry set
is>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>"+map_tclaim.entrySet());
m4.getKey();
String values = m4.getValue().toString();
String first=values.replace("[", "").replace("]", "");
String a[]=first.split("~~");
String part1=a[0];

System.out.println("Column 1
is>>>>>>>>>>>>>>>>>>>>"+part1);
String part2=a[1];
System.out.println("Column 2
is>>>>>>>>>>>>>>>>>>>>"+part2);
```

```
/*   part5= Temp.removehttplink(part5);*/

part2= Temp.removeUrl(part2);
System.out.println("String                                    part2
is>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>"+part2);
data=Preprocessing1.filter(part2);
StringBufferbuf = null;
String str=""+data;
data.add(str);
for(int ii=0;ii<data.size();ii++)
{
    buf=new StringBuffer();
    buf.append(data.get(ii));
}
String s=buf.toString();
s=s.replace("[", "");
s=s.replace("]", "");
ff=        AdminDAO.storefiltereddata(part1,
college,s);



}


booleandd=    Sentimentprocessexicution.senti(college);

System.out.println("ddddddddd"+dd);

if(dd)

{
String poss="positive";
    String negg="negative";
    String neuu="neutral";

    String undecidable="undecidable";
positivecount11 = PoolingDAO.getpositivecount(college,poss);
```

```
                    System.out.println("positivecount                    is
>>>>>>>>>>>>>>>>>>>>>"+positivecount11);

                    negativecount11 = PoolingDAO.getnegativecount(college,negg);
                        System.out.println("negativecount                    is
>>>>>>>>>>>>>>>>>>>>>"+negativecount11);

                        neutralcount11                                       =
PoolingDAO.getneuralcount(college,neuu);
                            System.out.println("nuetralcount                 is
>>>>>>>>>>>>>>>>>>>>>"+neutralcount11);


                            undecidablecount11                               =
PoolingDAO.getundecidablecount(college,undecidable);
                                System.out.println("undecidablecount         is
>>>>>>>>>>>>>>>>>>>>>"+undecidablecount11);

                            totalcount = PoolingDAO.gettotalcount(college);
                                System.out.println("totalcount               is
>>>>>>>>>>>>>>>>>>>>>"+totalcount);



                            finalcount=
AdminDAO.insertfinalcount(college,positivecount11,negativecount11,neutralcount11,unde
cidablecount11,totalcount);



            }}}

                    if(finalcount)
                    {
                        ResultSetrs=LoginDao.gettweets(college);
                        request.setAttribute("rs",rs);

                        RequestDispatcher
rd=request.getRequestDispatcher("/jsp/viewtsentimenttweets.jsp?no=2&pmark="+positive
```

```
count11+"&negmrk="+negativecount11+"&neumark="+neutralcount11+"&unmark="+un
decidablecount11+"&totcount="+totalcount+"");

                                rd.forward(request, response);



                        }


                }


                        catch (Exception e) {
                        // TODO: handle exception
                }


        }


}
```

LIST OF PUBLICATION

1. **Review paper:** International Journal of Information and Computer Science (IJICS), ISSN NO:9972-1347 Issue.25 Dec 2018.

2. **Research paper: "**Evaluation of product rating using data mining", International Journal of Information and Computer Science (IJICS), Vol. 06,Issue .05 May 2019
   ISO: 7021-2008 Certified journal

CURRICULUM VITAE

Name: Priti kumari

Postel Address: Near-R.S tank navtolia laheriasarai darbhanga bihar-846001

Mobile no: 8604809403

E-mail: kprit806@gmail.com

Nationality: Indian

Language known: English,Hindi

CAREER OBJECTIVE:

I am seeking a competitive and challenging environment where i can serve your organization and establish a career for myself.

TECHNICAL COMPETENCIES:

- C programmimg
- Java
- Unified modeling language

PERSONAL COMPETENCIES:

- Adaptive
- Ready to take challenge
- Self motivated
- Trustworthy
- Teamwork

ACADEMIC QUALIFICATION:

S.no	Degree	University/Board	Institute	Year	Percentage
1	High school	CBSE	Holy cross school	2010	64.6
2	Intermidiate	CBSE	Jesus and mary academy	2012	64.2
3	B tech(CSE)	AKTU	BBDEC	2016	64.9
4	M Tech	BBDU	BBDU	2019	79.9(1st yr)

PERSONAL PURSUIT:

- Painting
- Crafting
- Cooking

STRENGTH:

- Enthusiastic, creative and willing to assume increase responsibility.
- Ability to adapt quickly to challenges and changing environment.
- Positive attitute,strong willing,helpful in nature.

DECLARATION:

I hereby declare that above given information is true to my best of knowledge.

Annexure IV

BBD-PG-FORM 02

BABU BANARASI DAS UNIVERSITY, LUCKNOW

CERTIFICATE OF THESIS SUBMISSION FOR EVALUATION

1. Name: .....
2. Enrollment no. : .....
3. Thesis title: .....

   .....
4. Degree for which the thesis is submitted: .....
5. Faculty of the University to which the thesis is submitted

   .....
6. Thesis Prepration Guide was reffered to preparing the thesis. ☐ Yes ☐ No
7. Specifications regarding thesis format have been closely followed. ☐ Yes ☐ No
8. The content of the thesis have been organized based on the guidelines. ☐ Yes ☐ No
9. The thesis has been prepared without resorting to plagiarism. ☐ Yes ☐ No
10. All resourse used have been cited appropriately. ☐ Yes ☐ No

11. The thesis has been not submitted elsewhere for a degree. ☐Yes ☐No

12. Submitted 2 spiral bound copies pluse one CD. ☐Yes ☐No

(Signature of Candidate)

Name: .....

Roll No.: .....

Enrollment no.: .....