Customer Satisfaction Measurement using Sentiment Analysis

Shaha Al-Otaibi, Allulo Alnassar, Asma Alshahrani, Amany Al-Mubarak,
Sara Albugami, Nada Almutiri, Aisha Albugami

Information Systems Department
College of Computer and Information Sciences
Princess Nourah Bint Abdulrahman University
Riyadh, Saudi Arabia

Abstract—Besides the traditional methods of targeting customers, social media presents its own set of opportunities. While companies look for a simple way with a large number of responses, social media platforms like Twitter can allow them to do just that. For example, by creating a hashtag and prompting followers to tweet their answers to some question they can quickly get a large number of answers about a question while simultaneously engaging their customers. Additionally, consumers share their opinions about services and products in public and with their social circles. This valuable data can be used to support business decisions. However, it is huge amounts of unstructured data that is difficult to extract meaningful information out of them. Social Media Analytics is the field which makes insights out of social media data and analyzes its sentiment rather than just reading and counting text. In this article, we used Twitter data to get insight from public opinion hidden in data. The support vector machine algorithm is used to classify sentiment of tweets whether it is positive or negative and the unigram applied as a feature extraction method. The experiments were conducted using large set of training dataset and the algorithm achieved high accuracy around 87%.

Keywords—Social media analytics; sentiment; classification; support vector machine; unigram

I.   INTRODUCTION

In the few recent years, the social media platforms have been growing while people build a global communication network on the Internet via many social media applications. Daily a huge volume of media is created on the social networks. For example, in Twitter - one of the most popular social media application - there are over 500 million tweets or posts per day[1]. It is a revolution of how media is created and distributed by sharing, and realizing messages without any control. Social media has an important impact on the field of business, advertisement, and e-commerce as it explains consumer behavior and feedback about particular business proposals, services and products. Opinions and purchase decisions of the people and organizations are now affected and sometimes taken as a response to the content of social media before going to the market and actually test the product. In social media, all data from posts, comments and replies needs measuring results and concluding insights out of them rather than just reading the opinions of others, this is known as social media analytics. Social media analytics are the practice of gathering data from social media platforms and analyzing that data to make business decisions. The most common use of social media analytics is to mine customer sentiment in order to support marketing and customer service activities. The importance of social media analytics is intuitive and flexibly used by companies, organizations and individuals to know the insight of the market. It helps companies to know customers' viewpoints and their comments on the quality of the products and services to make successful business decisions. The typical objectives include increasing revenues, reducing customer service costs, getting feedback on products and services, as well as improving public opinion of a particular product or business division [1], [2].

To clarify the concept of social media analytics, we should present the problem from two viewpoints: the business problem and the technical issues. As business problem, the pre-sale means knowing the activity of the competitors in the market. Hence, companies need to know the right time to release their products or services in the market. Additionally, they need to check the state of the market if there is a product similar to its product or service that will be launched in the market and compare with each other, as well as determine what is the positive and negative about those products or services and try to improve it. Then, they will able to add a competitive advantage in their product or service. After-sale, the companies need to check the social media feedback and customers' opinions about the product or service, they want to know how many of the followers, interacting, re-tweet, fans and replies about company's account and products. Finally, it helps companies to understand the experiences of others with the product. Second, the technical issues related to the difficulty in extracting information and data about a particular product and deciding whether it is negative or positive of the marketing products. Moreover, the social media analytics require accessing the Internet and needs a large space to store the collected data for processing. They also need to filter and clean massive data, wide, variety, noise and unrelated data sources in social media content as well as, the extraction of keywords and show off all Hashtag that works mention to the product account and others [1].

In this article, we aim to support the organizations and individuals in decision-making through providing analysis of products information, customer's opinions, and the reviews of

---

[1] https://blog.twitter.com/

products in the social media. Indeed, the issues arises the need of providing such analytics include, knowing about the competitors from other companies as well as the need to solve the lack of means and tools to evaluate the products on the market. The proposed system will help companies and organizations to get benefit information about their products and services. It will save beneficiary's time and serve them by learning more about their products and stimulate the work of producing a lot of quantity of products through knowing the viewpoints of their customers as well as, the information about competitors' products. As initial step of this work, we will cover the textual data about products and services in Twitter and apply the characteristics of the intended users of the system, such as age, gender, education, number of followers, etc.

The article is organized as follows: Section II presents the background information and related work. Section III demonstrates the design issues and the implementation details related to analyzing social media contents. Then, the experimental evaluation standard is presented in Section IV. Finally, we conclude this work in Section V.

II. BACKGROUND INFORMATION AND RELATED WORK

This section starts by presenting the background information. Then, we review the literature that related to the social media analytics.

A. Background Information

Data is the currency of social media marketing and the understanding of social media analytics is essential for making data useful. Hence, the analytics allow marketers to identify sentiments and trends in order to better meet their customer's needs [2]. Facebook, Twitter, Pinterest and other social networks continue to spread a torrent of data, and organizations need to measure the business value. Now if customer wants to buy a product, he/she is no longer limited to asking his/her friends and families because there are many product reviews on the Internet which give opinions of existing users of the product. For a company, it may no longer be necessary to conduct surveys, organize focus groups or employ external consultants in order to find consumer opinions about its products and those of its competitors because the user-generated content on the Web can already give them such information [3]. Businesses often struggle to measure consumer interest and to determine what social data is actually useful for them to collect. By utilizing sentiment analytics complemented with human intelligence, companies can filter out noise and—with the help of machine-learning technology—identify the critical data that advances their business.

This section presents the social media analytics framework, techniques, types of audience, social media network choices, and features of media analytics tools.

1) Social media analytics Framework: The typical framework involves three-stage process: capture, understand, and present [4]. During the work of Chong et al., they develop CUP framework that add the identify stage to allow the

identification of posts/tweets prior to the capture stage [5]. This identification is done using keywords which are determined by users. These keywords are then used in the automated scripts query requests to social network's API, e.g. Twitter API, collects posts/tweets containing those keywords. Therefore, the steps include: the identify stage is the data accessing stage that involves identifying relevant keywords to use in collecting social media data. Then, the capture stage is the data cleaning step that involves obtaining relevant social media data by listening to various social media sources, archiving relevant data and extracting pertinent information, hence not all data captured will be useful. Next, the understand which is the data analysis stage that selects relevant data for modeling, removing noisy, low quality data, and employing various advanced data analytic methods to analyze the data retained and gain insights from it. Finally, the present is the data visualization stage that deals with displaying findings from the understand stage in a meaningful way [6].

2) Social Media Analytics Techniques: Many techniques can be used for social media analytics. First, the Supervised Classification, where the classification is the separation or ordering of objects into classes. Text classification is automatically assign the texts into the predefined categories. In this machine learning technique, the classifier learns how to classify the categories of documents based on the features extracted from the set of training data. The supervised classification includes: Support Vector Machine (SVM), Naïve Bayes, Neural Network, K-nearest Neighbor, and Decision tree [7], [8]. A detailed review of the above classifiers along with their advantages and disadvantages are explained in [8] and [9]. Typical text classification process has the following steps: collect data, normalize data, analyze the input data, train the algorithm, test the algorithm, and apply on the target data [9]. Second, Unsupervised Text Mining/Clustering: Text clustering is unsupervised learning, where no label or target value is given for the data. It is a method of gathering items or (documents) based on some similar characteristics among them. It performs categorization of data items exclusively based on similarity among them. Most clustering algorithms need to know the number of categories in advance. Some researchers use clustering instead of classification in topic detection because it hard to find data set for new topics [10].

3) Types of Audience: Twitter subscribers are older in age and count way more than Facebook's [10] so they are likely generated more trustworthy opinions. Also, people share their opinions publicly on Twitter unlike Facebook where social interactions are often private [11]. For these reasons, we selected Twitter as data source. However, follow/friend action in Twitter is not mutual like in Facebook so social circle of a user is not clear.

4) Features of Media Analytics Tools: Most of media analytics tools accomplish goals like:

- *Competitive benchmarking*: The ability to view profile and content information for other accounts like competitors.

- *Centralized analytics*: A single place to see and compare statistics and metrics for all (or most) of your social media accounts.

- *Influencer identification*: A list of the accounts or people that engage (share, comment, etc.) with your content most frequently.

- *Tracking of common social activities*: Tracking of customer service related interactions, or other common social network activities.

- *Dashboards*: Pre-made or custom dashboards so that you can easily keep tabs on the accounts, competitors, and metrics that matter the most to you.

- *Reporting*: Exportable reports and data often coupled with scheduling and email delivery [12].

5) Review of Data Analytics Systems: The social media analytics systems can be divided into two types: Platform tools and Cross-Platform tools. Platform tools are provided by the official social media networks such as Twitter, Facebook, YouTube, ... etc. while Cross-Platform are commercial tools allow the user to analyze different social networks types [10].

First, we will list the data analytics platform tools which are provided by the official social media networks:

- Twitter Analytics has a built-in analytics platform. It is available to individuals as well as businesses. Number of tweets, tweet impressions, visits to profile, mentions and followers are all tracked. There are monthly statistics on most popular tweets, mentions, and followers for that month. User can click on any Tweet to see the impressions, likes, retweet, and engagements, but no sentiment analysis is provided.

- Facebook Insights are available to any of the admins of company page once user has over 30 fans. It displays detailed metrics about posts and the engagement they earn. Audience analysis can help to understand who is engaging more, and includes demographic and locations breakdown. Engagement metrics can be seen for each of posts, helping user to understand what type of content works best. There are also metrics on video views, actions taken on user page and the reach of your posts. Also here, no sentiment analysis is provided.

- YouTube Analytics provides an in-house analytics tool so anyone who has uploaded videos can understand their performance. The tool displays performance metrics, engagement metrics, and demographics. It helps user understands how people found videos, how much they watched, if they clicked through to user's website, and who they were.

- Google Alerts/Analytics While not strictly a social media analytics tool. It allows user to monitor the web for new content, mentions of brand, competitors, or industry thought leaders. Creating an Alert means you will receive email notifications when Google finds new results on the topic across blogs, forums and news sites. Google analytics is primarily a web analytics tool, but it provides a small but important role in social media analysis: a breakdown of which social sites are driving traffic to user's website.

Second, the Cross-platforms which are commercial tools allow the user to analyze different social networks types.

- SimplyMeasured is a paid tool, but has various free individual reports for Instagram, Twitter, Facebook, Vine and Google+. A follow of the SimplyMeasured Twitter account is enough to secure user's report. They provide insights such as Facebook content, competitor and fan page analysis, key Twitter analytics, Instagram engagement, content and trends analysis and many others.

- Quintly covers Facebook, Twitter, Google+, LinkedIn, Instagram and YouTube, and it has a free tool for Facebook analytics. Quintly comes with a standard dashboard that can be customized with widgets to suit user's needs and track the metrics that matter to user.

- Brandwatch crawls millions of sites and allows user to build flexible and accurate searches through advanced Boolean queries. Brandwatch categories, rules, and tags allow users to slice and dice the data any way they want.

B. Related Work

There are many surveys for data analytics and related topics, some of them will be presented in this paragraph. Bo and Lillian presented a survey that covered the techniques and approaches for opinion mining and sentiment analysis to promise enabling opinion-oriented information seeking systems. It provides a discussion of available resources, benchmark datasets, and evaluation campaigns were provided [13]. Isaac presented a survey of different social network analysis techniques employed in many applications interpreting social media data, e.g. Twitter. It focuses on two main approaches to sentiment analysis: supervised learning and unsupervised learning techniques used for natural language processing, classification and prediction. Major statistical packages such as SAS and SPSS include dedicated sentiment analysis modules used in [10]. Additionally, a review of text classification on social media data is to discuss the different types of classifiers and their advantages and disadvantages [8]. Moreover, a comparison of the most popular packages, e.g. R, Matlab, SciPy, Excel, SAS, SPSS, and Stata that are typically used for data analysis was presented in [14]. A book was published of mining data from the social web such as Facebook, Twitter, LinkedIn, Google+, GitHub and More. This book provided an explanation on how to acquire, analyze, and summarize data from social media networks, email, websites, and blogs by employing the Natural Language Toolkit, NetworkX, and other scientific computing tools to mine popular social web [15]. Moreover, Twitter data analytics book presented an understanding of the basics of collecting, storing, and analyzing Twitter data. The

first half of this book discusses collection and storage of data. The second half is focused on analysis. It provided the common measures and algorithms that are used to analyze social media data [16]. Finally, the text mining and analysis book covered the practical methods, examples, and case studies using SAS. It delivered a comprehensive theoretical reference for text mining as well as many practical examples and case studies using the Statistical Analysis System (SAS) [17].

There are many datasets used for data analytics provided in the literature, for example: The datasets of customer reviews, pros and cons as well as comparative opinions [18]. The MPQA opinion corpus provided opinion datasets, e.g. Subjectivity Lexicon [19]. Additionally, Twitter sentiment analysis training data contains corpus of already classified tweets in terms of sentiment analysis training and testing where it contains more than 1,500,000 classified tweets, each row is marked as "1" for positive sentiment and "0" for negative sentiment [20]. Moreover, Sanders-Twitter sentiment corpus designed for training and testing Twitter sentiment analysis algorithms. It consists of 5513 hand-classified tweets. These tweets were classified with respect to one of four different topics [21].

In this paragraph, we will present some of researches for data analytics tools. First, the sentiment analysis and text mining for social media microblogs using open source tools. It presents an empirical study that used R package to perform text mining and sentiment analysis for Twitter online reviews about two retail stores in UK [6]. Second is the experiment on

binary classification for Twitter sentiment analysis. This experiment demonstrates how to use Microsoft Azure Machine Learning Studio to train a text sentiment classification engine using the Two-Class SVM [22]. Third, the emotion classification of social media posts for estimating people's reactions to communicate alert messages during crises. This article describes a methodology for analyzing tweets about Sandy hurricane and annotating them with four emotional labels. Two classification algorithms were experimented: Naïve Bayes and SVM classifiers. The results show that the algorithm achieves the best results with about 60% accuracy [23]. Fourth, the localized Twitter opinion mining using sentiment analysis analyzes tweets about iPhone 6 using SentiWordNet, part of SNLP which is an open source natural language processing tool developed by Stanford University [24]. Finally, the data mining and analysis on Twitter++ study start with a few discussions of how geo-tagged tweets in Twitter can be used to identify useful user features and behaviors as well as identify places of interests. Then, it presents a clustering analysis and proposes different similarity measures to detect communities [25].

Many tutorials describe how to analysis Twitter data, for example, step-by-step practical tutorials build Twitter analytics tool with R package included in [26]-[28]. Additionally, the tutorials designed to build Twitter mining tool with Python are included in [29] and [12]. Finally, the practical tutorials build Twitter mining tool with MATLAB are included in [11] and [30]. Table I illustrates a comparative analysis of some presented works.

TABLE I.   A COMPARATIVE ANALYSIS OF SIMILAR WORKS

Ref.	Sentiment Analysis	Network Analysis	Customer Service	Application Data	Social Media Network	Analysis Method	Programming Package
[6]	Yes	No	No	Customer Review	Twitter (Facebook applicable)	Data Mining (association rules) (Lexicon-based)	R Package (Twitter)
[22]	Yes	No	No	Sentiment140 dataset	Twitter	Classification (SVM) 2 classes	Microsoft Azure Machine Learning Studio + R Package
[23]	Yes	No	No	Sandy hurricane	Twitter	Classification (SVM) 4 classes	Python Package (tweetstream)
[24]	Yes	No	No	Reviews on iPhone6	Twitter	Natural Language Processing	Stanford NLP (SentiWordNet)
[25]	Yes	Yes	No	geo-tagged	Twitter	Data Mining (Clustering)	MATLAB (Twitter 4j)
[33]	Yes	Yes	No	Slahdot Lexicon (MPQA)	Slashdot Website (comments on news)	Data Mining Predictive Analytic	KNIME

III.   PROPOSED SOLUTION OF DATA ANALYTICS

Our proposed solution started by data collection which is an important aspect of any type of research study. Hence, the choice of data collection method is influenced by the data collection strategy, the type of variable, the accuracy required, the collection point and the skill of the source. The main data collection methods we used: first, the literature review and

tools analysis. It supported us for collecting set of requirements regarding the analysis algorithm, analysis metrics as well as user interface design.   Second, set of interviews were conducted with the respondent and notes are subsequently interpreted for further analysis. We conducted set of interviews with clients selling their home-made products, such as accessories and crafts, using different social media networks. Answers from respondents mainly raised the issue that searching within social media is very difficult for

them to target particular categories with people such as customers/competitors existing in particular country, are of particular age, females as well as customers which are influencers and having high number of followers. Targeting the right customers and monitoring the right competitors will bring them higher profits. Hence, we included one requirement about filtering input data against the criteria they mentioned by them. Third, we used questionnaires which are completed and returned by respondents. We have used GoogleForms to design a questionnaire that contains 16 questions of many types (yes/no, multiple-choice and open answer) and directed to different categories of people, i.e. students, tutors, business owners and consumers. The result of questioner let us focus on analyzing Twitter data since it will be more useful to target large number of people. We plan to satisfy the following SW/HW requirements:

- Libraries to communicate with Twitter API to authenticate added Twitter accounts and retrieve of Twitter data.

- Statistical and Machine learning development packages such as LibSVM, WEKA or R.

- Benchmark tweets database for customer reviews on an arbitrary product.

- Lexicon dictionary of sentiment words classified as positive and negative.

- Laptop machine with at least 8 GB of RAM and no less than Terabyte disk.

- Public server with high quality feature to upload the system and accommodate huge amounts of data.

Moreover, the nonfunctional requirements that should be satisfied are:

- Security/privacy by providing access permissions for system data, i.e. login/logout, valid emails and authorized Twitter accounts.

- Availability: The system available for service when requested by users.

- Usability: Simple UI to provide easy-to-learn end system.

- Reliability: The ability of a system to perform its required functions with accuracy no less than 80%.

- Visualization: The system should display metrics visually as well as numerically. Visual presentation includes keyword cloud, bar charts, pie charts, trend graphs and comparative graphs while numerical includes totals and percentages in addition to specific scores.

- Sentiment analyzed tweets are marked in different colors for negative and positive.

A. System Design of Proposed Solution

This section illustrates the design of the proposed solution and its architecture including the structure, the description

about each structure component and the used system design tools. The system will be implemented in five-tier server-client architecture model consisting of presentation layer, business logic layer and data access layer for internal components. However, the additional integration layer and data Source layer are used to describe external components. Fig. 1 illustrates the main system's architecture and components.
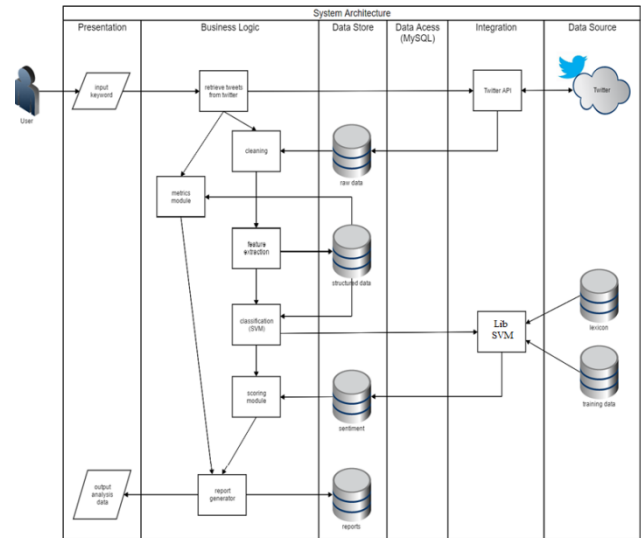

Fig. 1. System architecture.

Presentation Tier, this layer contains the user oriented functionality responsible for managing user interaction with the system, and generally consists of components that provide a common bridge into the core business logic encapsulated in the business layer [31]. In the proposed system, the presentation layer does two tasks: accepts user's input data such as keyword list and the type of analysis report and the other task is to later visualize the analysis results.

Business Logic Tier, it implements business functionality of the system. For example, it moves and processes data between the two surrounding layers. In our proposed system, the business logic layer consists of the following tasks:

1) **Tweets retrieval,** Twitter is the most exaggerated part of social networking site, it consists of various blogs which are related to various topics worldwide. Instead of taking whole tweets, we will rather search on particular keywords and store all tweets in the form of text files by using mining tool i.e. WEKA/R/LibSVM which provides sentiment classifiers.

2) **Cleaning and Pre-processing of extracted data,** after retrieval of tweets, sentiment analysis tool is applied on raw of tweets but in most of cases, it gives poor performance. Therefore, preprocessing techniques are necessary for obtaining better results. We extract tweets, i.e. short messages from Twitter which are used as raw data. This raw data needs to be preprocessed. So, preprocessing involves following steps:

- Exclude tweets with non-English languages.

- Remove emoticons and substitute with their textual meanings.

- Remove URL's, hashtag mark "#", mentions "@" and retweet prefix, i.e. "RT".

- Remove punctuation marks and articles such as "a", "an" and "the".

- Remove stop words such as "to", "of", "is", "are", "this" and "for".

- Normalize elongated words, e.g., happyyyyyy, by only one or two occurrences only.

3) **Feature extraction,** in feature extraction method, we extract the aspects from the processed dataset. Later this aspect is used to compute the positive and negative polarity in a sentence which is useful for determining the opinion of the individuals using models like unigram and bigram. Additionally, the machine learning techniques require representing the key features of text or documents for processing. These key features are handled as feature vectors which are used for the classification task. The features extraction method that will be considered in this system is unigram.

4) **Sentiment Classification,** the selected sentiment classifier is the SVM as it scores higher than other approaches according to [6]. Training of classifier data is the main motive of this step. A reference model is derived based on the analysis of a set of training data. Training data consists of data objects whose class labels are known. The derived model can be represented in various forms, such as classification (IF-THEN) rules, decision trees, mathematical formulae, or neural networks. Classification process is done in a two-step process as illustrates in Fig. 2. First step is Training in which we will build a model from the training set. Second step is Prediction in which we will check the accuracy of the model and use it for classifying new data

5) **Sentiment Scoring Module,** we use the lexicon/dictionary that applied in [6] in which English language words assigns a score to every word, between 1 (Negative) to 3 (Positive). So, this scoring module is going to determine score of sentiments in the sentiment analysis of data. Based on the dictionary assignment of score, the system interprets whether the tweet is positive, negative or neutral.

6) **Computing metrics,** this component is irrelevant to sentiments, however, it computes meaningful measurements about tweets and Twitter users. The raw data that comes from Twitter API contains the following parameters of each tweet which later can be used to calculate metrics:

- **Likes:** list of people liked this tweet. It's usually positive in sentiment.

- **Followers**: list of people that are currently subscribed to this tweet.

- **Mentions**: list of @username included in this tweet.

- **Replies**: list of responses to this tweet that begins with tweet writer's @username.

- **Retweet (RT):** list of users who shared this tweet.

Data Access Tier, the data Access Tier communicates with the database. In the proposed system, we are going to use MySQL DBMS to manage data storage, querying and retrieval.

Integration Tier, this tier is responsible for communicating with external resources and systems such as data stores, API's and legacy applications. The business tier is coupled with the integration tier whenever the business objects require data or services that reside in the resource tier. The components in this tier can use some proprietary middleware to work with the resource tier [4]. In the proposed system, this layer contains components interacting with Twitter API in order to access Twitter data in addition to open source library such as WEKA or R in order to use their functions and classes implementing machine learning algorithms, e.g. SVM.

Data Source Tier, this tier contains the business data and external resources such as Twitter network, training data source and lexicon benchmark.
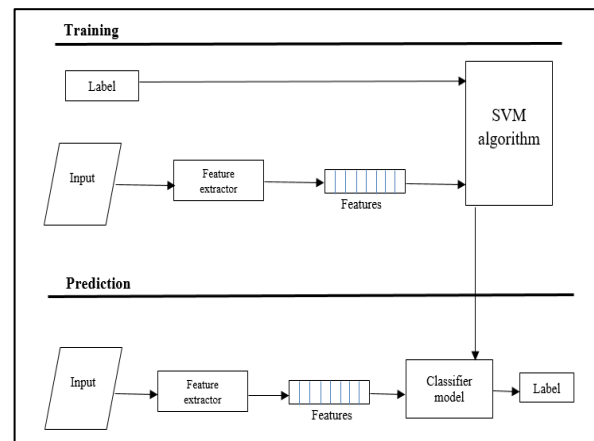

Fig. 2. Training SVM vs. Prediction.

B. System Implementation

In this phase, we take the determined system specifications and code them. The implementation requirements include the following software and hardware specification.

1) Hardware Requirements
- Laptop with processor Intel core i5, minimum speed of 1.7 GHz and 8.00 GB of RAM for faster running and better performance.

- Internet connection to access Twitter network and bring Twitter data online.

2) Software Requirements
- Window 10, 64-bit operating system or similar alternatives.

- As for user interface design, we used a bootstrap HTML5 free template called SIMINTA as a ground for our design.

- The website of proposed system was implemented using HTML5/CSS3 for page design, PHP for server-side scripts. Therefore, an Apache server distribution, such as XAMPP, was needed to execute PHP scripts.

- XAMPP also includes MySQL server which we used to store users' accounts and analysis reports' data. The database was managed using PHPMyAdmin module in XAMPP.

- Registered our application on Twitter Application Management to get Twitter access tokens and authorization. For the implementation of Twitter API interface, the Twitter -API-PHP library applied while it was recommended by Twitter developers' page.

- In addition, we used an executable software to run SVM classification algorithm from LIBSVM, which is a library developed for Support Vector Machines.

- For preprocessing, algorithm training and testing, we utilize different training datasets and some lingual dictionaries including stop word list, acronyms dictionary, positive and negative tweets provided by [16].

- To draw charts, we use classes from PHPLOT free php library.

3) System Major Services

The proposed system provides the following major services to its users:

- Account Analysis: user can search for specific Twitter account and analyze its author's activity rate in addition to followers' engagement with this account for the last ten days. For example, user can monitor his own product's account or a competitor's public account.

- Keyword Analysis: user can search Twitter social media network for any keyword, hashtag or mention of interest to check the public opinion and other valuable indicators about it. Keyword Analysis is on three types:

   a) Sentiment Analysis: The percentage of latest positive vs. negative tweets talked about this search term. Search term can be a company name or a product brand for example.

   b) Compute Metrics: The strength and reach of this search term in the public. Top hashtags and top keywords accompanied with this search term as well as and top Twitter users who are most interested about this search term.

   c) Comparative Analysis: Providing sentiment analysis and metrics for two opposed search terms.

- Reports: Results of keyword analysis including individual analysis and comparative analysis can be stored in the database and retrieved back as needed. It can also be printed out or saved as PDF.

4) Implementation Details

In this section, we will give a brief description of how the proposed system was actually implemented using the specified software and hardware requirements. Firstly, we mention that our proposed website named TweetAdvisor.

- **System Website Registration in Twitter Apps:** Twitter, as many other social networks, have its own web services API (Application Programming Interface) that applications, such as our website, can work with. However, in order to use Twitter web services API, the first step is to register our website on Twitter's Application Management. After that, it will be provided the necessary access and authentication tokens to access Twitter data and services.

- **Twitter REST API:** After we registered the system's website to Twitter apps world, we need to access and call the appropriate Twitter 's web services to handle the website functions. The REST APIs provide programmatic access to read and write Twitter data. Read user profile, timeline or search Twitter data, and more. The REST API identifies Twitter applications and users using OAuth; responses are in JSON format. Basically, we used the following three GET web services from Twitter:

   *a).*GET search/tweets: returns a collection of Tweets matching a query.

   *b).*GET users/show: returns profile information about user specified by the user_id or screen_name parameter in the query.

   *c).*GET statuses/user_timeline: returns a collection of the most recent Tweets posted by the user indicated by the screen_name or user_id parameters in the query.

- **Sign Up/Sign In:** The importance of creating an account is that private analysis reports conducted by the user can be saved and retrieved. User can create an account that providing basic personal information such as username, email and password.

- **Keyword** *Analysis: Tweets Fetch*: User can search Twitter social media network for any keyword, hashtag or mention of interest to check the public opinion and other valuable indicators about it as illustrates in Fig. 3. Keyword Analysis is on three types Sentiment Analysis, Compute Metrics and Comparative Analysis. Additionally, the proposed system allows the user to determine the following parameters:

   a) **Exclude**: it returns all the tweets that don't contain specified words/phrases.

   b) **From**: it retunes all the tweets coming from the specified user's screen name.

   c) **min_followers_count**: it returns all the tweets only written by users who have a minimum number of followers, i.e. target influencers or famous users.

After specifying the required parameters, a query will be sent to Twitter API in order to retrieve tweets result. The request to Twitter web service is accepted only if the authentication via access tokens passed. Access tokens are given after successful app registration as explained in previous section. The search query will return the result data in JSON

tree format which is converted into an array object and then saved in PHP session for the next step; the preprocessing. Before preprocessing, the raw result array is filtered to only include tweets which are more than 20 characters in length and exclude retweets and redundant tweets as appears in Fig. 4. For connection with Twitter API we use a PHP Twitter -API-PHP library recommended by Twitter developers' page.
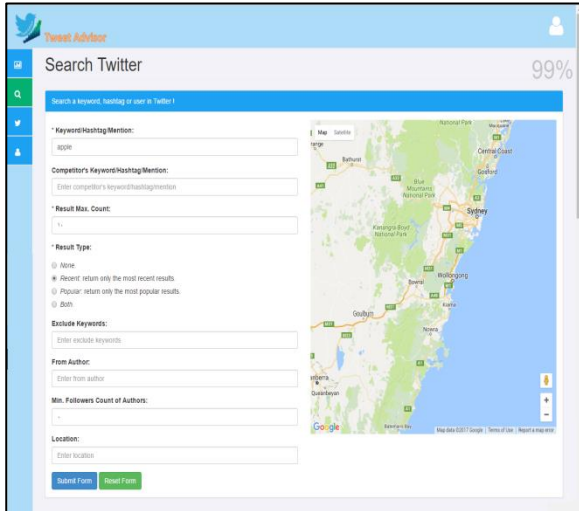

Fig. 3.   Search criteria in Twitter.

- **Keyword Analysis: Preprocessing:** Preprocessing is the step needed to clean the data from noise, standardize and convert it to a structured format before extracting distinct features from it. In this work, we used four external resources in order to preprocess the data and provide prior score for some of the commonly used words:
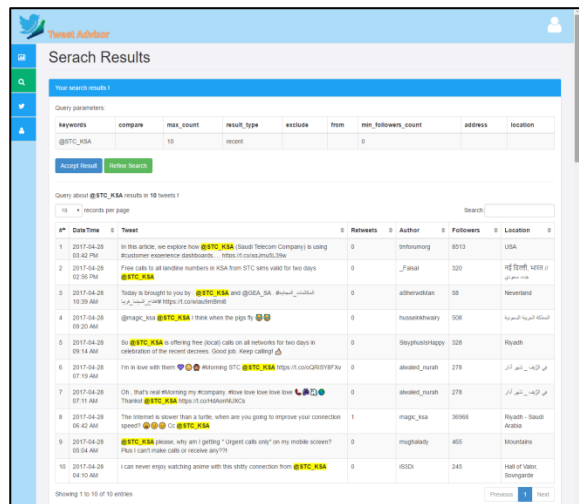

Fig. 4.   Search results screen.

a) **Emoticon Dictionary**: Each emoticon is annotated into one of four classes: a) Extremely-Positive; b) Positive; c) Extremely-Negative; d) Negative as given in [32].

b) **Acronym Dictionary**: We used the acronym expansion list as given in [32].

c) **SentiWord List**: is a list of English words classified by its POS (short for position in sentence) and rated for valence with an integer between minus eight (negative) and plus eight (positive). POS types are noun (N), verb (V), adjective (A) and adverb (R). We used the SentiWord list as given in [32].

d) **Stop Words**: is a list such as a, is, the, with, and, or, I, you, etc. which occur in high frequency in a sentence but don't carry any sentiment information and thus are of no use to us. We used the stop words list as given in [32].

After building these lingual and sentiment dictionaries, preprocessing of tweets starts. We following preprocessing steps: remove extra whitespaces, replace each acronym with its expansion, tokenize each tweet, i.e. split into an array of separate words. then for each word in a tweet do the following, remove 'RT' prefix, lowercase , replace url, with ||U||, replace hashtag sign, i.e. #, with a ||H||, replace exclamation mark with ||EXC||, user mention replaced it with ||T||, remove punctuations, remove all digits, if emoticon, replace it with its equivalent sentiment into one of ||N|| for 'Negative', ||XN|| for 'Extremely-Negative', ||P|| for 'Positive'; or ||XP|| for 'Extremely-Positive', replace each "n't", "no", "not", "never", "cannot" with ||NOT|, if a stop word, remove it, replace tag word with its position of sentence + ||POS|| + word. An example of a preprocessed tweet is given in Table II.

After preprocessing, tweets are stored in array session so feature extraction phase starts.

Keyword Analysis: Unigram Feature Extraction: Feature vector is the most important concept in implementing a classifier. A good feature vector directly determines how successful the classifier will be. The feature vector is used to build a model which the classifier learns from the training data and further can be used to classify previously unseen data. In tweets training data, consisting of positive and negative tweets, we can split each tweet into words and add each word to the feature bag. Adding individual (single) words to the feature bag is referred to as 'unigrams' approach, see Table III.

So in unigram features, each feature is a single word found in a tweet. If the feature is present, the value is 1, but if the feature is absent, then the value of this feature is just not included. The entire feature vector of each tweet will be a combination of each of these feature words and based on this pattern, a tweet is labeled as positive or negative. See, Tables IV and V.

TABLE II.   AN EXAMPLE OF A PREPROCESSED TWEET

Tweet	#Wearables | Apple Music reportedly strikes deal with popular social media app Musical.ly :D :D https://t.co/mobfOW1YIS… https://t.co/9bJGxs986r
Pre-processed	||H||wearables N||POS||apple N||POS||music R||POS||reportedly V||POS||strikes V||POS||deal A||POS||popular A||POS||social N||POS||media N||POS||application ||P|| ||P|| A||POS||musical N||POS||love ||U|| ||U||

TABLE III. UNIGRAM APPROACH – BAG OF WORDS

Tweet (Positive)	#Wearables | Apple Music reportedly strikes deal with popular social media app Musical.ly: D: D https://t.co/mobfOW1YIS… https://t.co/9bJGxs986r
Pre-processed	||H||wearables N||POS||apple N||POS||music R||POS||reportedly V||POS||strikes V||POS||deal A||POS||popular A||POS||social N||POS||media N||POS||application ||P|| ||P|| A||POS||musical N||POS||love ||U|| ||U||
Unigram Feature Bag	{ wearables, apple, music, reportedly, strikes, popular, social, media, application, love }
Tweet (Negative)	My brother lost his phone in his room and my mom calling me trynna get me to do the find my phone shit. 🙁
Pre-processed	||N||POS||brother V||POS||lost N||POS||phone N||POS||room N||POS||mom V||POS||callingV||POS||get V||POS||do V||POS||find N||POS||phone N||POS||shit ||N||
Unigram Feature Bag	{ wearables, apple, music, reportedly, strikes , popular, social, media, application, love, brother, lost, phone, room, mom, calling, get, do, find, shit }

Some of the other feature vectors also add 'bi-grams' in combination with 'unigrams'. For example, 'not good' (bigram) completely changes the sentiment compared to adding 'not' and 'good' individually. Here, for simplicity, we will only consider the unigrams.

TABLE IV. UNIGRAM FEATURE VECTORS

Tweet 1 (Positive)	||H||wearables N||POS||apple N||POS||music R||POS||reportedly V||POS||strikes V||POS||deal A||POS||popular A||POS||social N||POS||media N||POS||application ||P|| ||P|| A||POS||musical N||POS||love ||U|| ||U||		
Tweet 2 (Negative)	||N||POS||brother V||POS||lost N||POS||phone N||POS||room N||POS||mom V||POS||calling V||POS||get V||POS||do V||POS||find N||POS||iphone N||POS||shit ||N||		
Unigram Feature Bag	{ wearables, apple, music, reportedly, strikes , popular, social, media, application, love, brother, lost, phone, room, mom, calling, get, do, find, shit }	Tweet 1 1:1, 2:1, 3:1, 4:1, 5:1, 6:1, 7:1, 8:1, 9:1, 10:1	Tweet 2 11:1, 12:1, 13:1, 14:1, 15:1, 16:1, 17:1, 18:1, 19:1, 20:1

TABLE V. FEATURE VECTORS FILE

+1 1:1, 2:1, 3:1, 4:1, 5:1, 6:1, 7:1, 8:1, 9:1, 10:1 // *Tweet 1*
-1 11:1, 12:1, 13:1, 14:1, 15:1, 16:1, 17:1, 18:1, 19:1, 20:1 // *Tweet 2*

- **Keyword Analysis: SVM-Classification:** The feature extraction method contains both training and testing. In training, the tweets vectors are labeled with '+1' for positive and '-1' for negative as in Table V. The classifier will use labeled vectors to learn from them and builds its learning model. In testing, however, each new un-labeled tweet will be compared to the bag of word generated from labeled tweets to create the new vector in the same way, however, with no labels given. The classifier will take the model and the new un-labeled vectors to predict the new classification results. The used classifier is based on SVM algorithm and provided by LibSVM library as two main executable applications: *svm-predict.exe* and *svm-train.exe* [6]. Fig. 5 illustrates the SVM classification result.
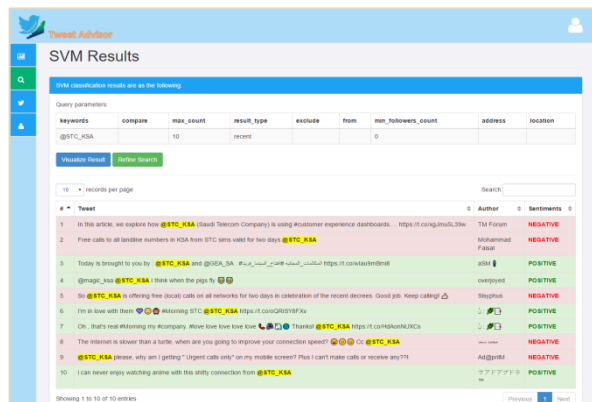

Fig. 5. SVM classification results screen.

- **Keyword Analysis: Compute Metrics:** After SVM classification completes, the following list of metrics will be calculated:

a) **Positive vs. Negative** counts pie chart: It indicates the percentage of passivity and negativity of public opinion about this search term.

b) **Strength:** It is the percentage of tweets count in last 24 hours on the total count of tweets in the result. It indicates how recent this search term is.

c) **Reach:** It is the percentage of different authors' count on the total count of tweets in the result. It indicates the percentage of authors interested in and talking about this search term.

d) **Top Keyword:** It is the list of the most frequented six keywords in the results. It indicated what other topics are related to this search term.

e) **Top Hashtags:** It is the list of the most frequented six hashtags in the results. It indicated what hashtags are related to this search term.

f) **Top Authors:** It is the list of the most Twitter accounts talked about the search term. If reach is 100% then each author has exactly one tweet in total result.

Fig. 6 visualizes a graphical result that illustrates the previous calculated metrics.

- **Keyword Analysis: Comparative Analysis:** We implemented this function by maintaining an array of

keywords entered by the user in the search query. Then, for each step, we run the code in a loop of the size of this array, and store each associated data results list in the corresponding array element. In this way, we will end up with multiple results each stored in its own array as appears in Fig. 7.
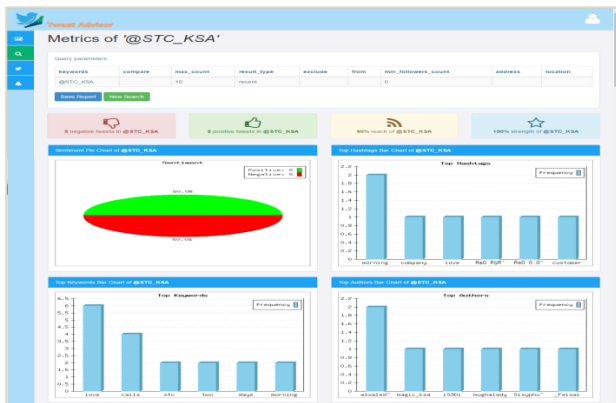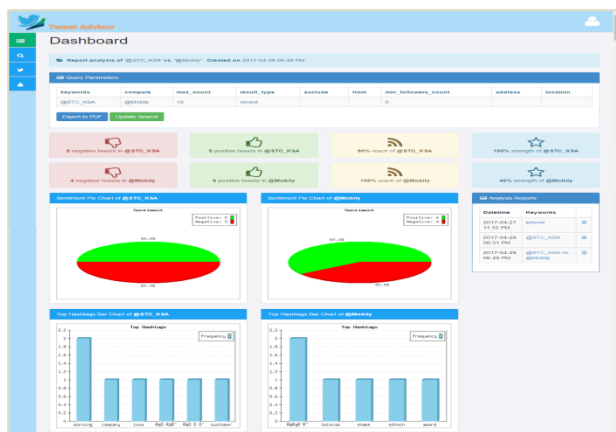

Fig. 6. Visualize metrics screen.

Fig. 7. Comparative keywords analysis.

- **Account Analysis:** User can search for specific Twitter account and analyze its author's activity rate in addition to followers' engagement with this account for the last ten days. For example, user can monitor his own product's account or a competitor's public account. Fig. 8 displays an account analysis for STC company. To implement this function, we need to send queries to *GET users/show* to retrieve user account's information such as *followers_count*. Also, we need to request *GET statuses/user_timeline* to extract tweets posted for last 10 days and get *favourite_count* and *retweet_count* for each tweet. Out of this data, we can calculate the following analysis metrics: Followers count so far, Tweets count last 10 days, Daily tweeting average, Daily interactions with followers, i.e. reply, average, Total likes by followers last 10 days, Average likes by followers per tweet, Total retweets by followers last 10 days, and average retweets by followers per tweet. Moreover, we plot the following graphs:

- **Followers Engagement Graph**: It represents the total likes and retweets by followers for each day during the interval of last ten days. User can notice at what day the engagement was higher or lower and analyze this reading.

- **Author's Activity Graph**: It represents the total tweets, retweets and interactions, i.e. replies, made by the owner of the account for each day during the interval of last ten days. User can notice at what day the activity was higher or lower and analyze this reading.

Fig. 8. Account analysis screen.

IV. EXPERIMENTAL EVALUATION STANDARD

In order to evaluate the efficiency of TweetAdvisor, we conduct a testing process to evaluate the system and its components with the intent to find whether it satisfies the specified requirements or not. In this work, we have considered and performed the following testing types: we first test the used classification algorithm SVM which is a Machine learning method that usually have to deal with big and uncertain data, and the output of the system which is not like traditional system having a good sign of right or wrong.

Therefore, to test a machine learning algorithm accuracy, we need a training dataset, a testing dataset and an independent piece of code as a benchmark to run the algorithm and check the accuracy results. Sometimes, it is better to use different datasets and analyze the properties of the dataset that improved the accuracy of the algorithm.

A. Test Results

In this section, we will present the testing results of SVM classification process in detail as it is the core functionality of our website. We fetched 6 tweets about search term "Google Chrome" and result type = "Both: recent and popular", see Tables VI, VII, VIII and IX.

TABLE VI.   RAW TWEETS RESULT

1	Save some valuable time with 20 of the best Google Chrome extensions for social media marketers: https://t.co/94L8Fs7IzV
2	Alert! Google Chrome Is Listening In To Your Room, Sending Data Without Consent - YouTube https://t.co/7QFCfoQYug April 29, 2017 at 06:00AM
3	Big UX step forward: starting in Chrome 59, web notifications will be shown directly as Mac OS X notifications. https://t.co/DyIHjoyjn1
4	Finally! Chrome will soon start using native notifications on macOS. https://t.co/pMCiu94jAj
5	13 Google Chrome Life Hacks that will Blow Your Mind! https://t.co/emjwu6VEuM
6	Google gets more aggressive in alerting users when web pages are not secure https://t.co/VaqjPYR5ID

TABLE VII.   PREPROCESSED TWEETS RESULT

1	V||POS||save A||POS||valuable N||POS||time A||POS||best N||POS||google N||POS||chrome N||POS||extensions A||POS||social N||POS||media N||POS||marketers ||U||
2	V||POS||alert ||EXC|| N||POS||google N||POS||chrome V||POS||listening N||POS||room V||POS||sending N||POS||Data N||POS||consent youtube ||U||
3	A||POS||big N||POS||user N||POS||experience N||POS||step A||POS||forward A||POS||starting N||POS||chrome N||POS||web N||POS||notifications R||POS||directly N||POS||mac A||POS||operating N||POS||system N||POS||notifications ||U||
4	R||POS||finally ||EXC|| N||POS||chrome V||POS||start V||POS||using A||POS||native N||POS||notifications macOS ||U||
5	N||POS||google N||POS||chrome N||POS||life N||POS||hacks V||POS||blow N||POS||mind ||EXC|| ||U||
6	N||POS||google A||POS||aggressive N||POS||alerting N||POS||users N||POS||web ||NOT|| A||POS||secure ||U||

Finally, the testing results show that the accuracy given by testing the test cases reached 84%, i.e. one error out of six correct answers.

B. Training/Testing Dataset Collections

The second test was performed using the publicly available data sets of Twitter messages with sentiment analysis. We have used a combination of two datasets to train the SVM machine learning classifier. For the test dataset, we randomly choose 4000 tweets which were not used to train the classifier. The details of the training and test data are explained in Table X.

TABLE VIII.   UNIGRAM FEATURE VECTORS RESULT

1	284:1 318:1 756:1 993:1 1004:1 1088:1 1381:1 1496:1 1554:1 2457:1 2766:1 3050:1 3903:1 4069:1 4636:1 5089:1 6553:1 7618:1 8136:1 8825:1 9017:1 9042:1 9090:1 9384:1 9564:1 9707:1 9729:1 10616:1 10783:1 10934:1 11134:1 11926:1 13071:1 14282:1 15093:1 15094:1 15691:1 17041:1 17567:1 17570:1 17629:1 17630:1 17631:1
2	121:1 404:1 711:1 1381:1 1496:1 2457:1 3659:1 3903:1 4274:1 4701:1 5610:1 5994:1 6021:1 6119:1 6744:1 7618:1 8136:1 8186:1 9017:1 9042:1 9115:1 9564:1 9729:1 10029:1 10783:1 10933:1 11363:1 11554:1 11704:1 12477:1 12929:1 13042:1 15093:1 15094:1 15691:1 15913:1 16384:1 16421:1 16489:1 16557:1 17041:1 17629:1 17630:1 17741:1
3	45:1 86:1 317:1 885:1 1381:1 1496:1 1717:1 1770:1 2242:1 2457:1 2854:1 3543:1 3616:1 3747:1 3903:1 4992:1 5014:1 5483:1 5727:1 5993:1 6017:1 6744:1 6846:1 7105:1 7618:1 7929:1 8100:1 8136:1 9042:1 9159:1 9405:1 9564:1 9729:1 10110:1 10354:1 11554:1 11576:1 11774:1 12339:1 15093:1 15094:1 15691:1 17147:1 17629:1 17630:1 17741:1 17758:1
4	1661:1 2457:1 3478:1 3659:1 3747:1 3823:1 5727:1 6017:1 6569:1 6744:1 6974:1 7403:1 7618:1 8136:1 8507:1 8825:1 9042:1 9405:1 9564:1 9729:1 9730:1 10934:1 11554:1 11774:1 13052:1 14017:1 15093:1 15691:1 17629:1 17630:1 17741:1 17758:1
5	98:1 107:1 630:1 1381:1 1496:1 2457:1 3903:1 5427:1 7167:1 7618:1 8136:1 9017:1 9042:1 9564:1 11554:1 11774:1 15093:1 15094:1 15691:1 17041:1 17629:1 17630:1 17758:1
6	1381:1 1496:1 2216:1 2263:1 2457:1 5993:1 8186:1 8825:1 9017:1 9042:1 9152:1 9273:1 9564:1 9729:1 10110:1 10403:1 10426:1 10934:1 11554:1 12339:1 15093:1 15094:1 17041:1 17567:1 17629:1 17741:1

TABLE IX.   SVM CLASSIFICATION RESULT

#	SVM Classification	Meaning	Human Classification	Meaning
1	-1	negative	1	positive
2	-1	negative	-1	negative
3	1	positive	1	positive
4	1	positive	1	positive
5	1	positive	1	positive
6	-1	negative	-1	negative

Sanders corpus is designed for training and testing Twitter sentiment analysis algorithms. It consists of 5513 hand-classified tweets. These tweets were classified with respect to one of four different topics. Each entry contains: Tweet id, Tweet text, Tweet creation date, Topic used for sentiment, and Sentiment label i.e. 'positive', 'neutral', 'negative', or 'irrelevant'. We used only the positive and negative tweets out of this dataset for training. To fetch random testing tweets, we used our website interface which searches the Twitter API for a given keyword with recent results. Tweets were downloaded, manually labeled and then subjected to both preprocessing and feature extraction as specified in Section 3. These filtered tweets are fed into the trained classifiers and the resulting output is then saved in a file. The results file was read and compared with the correct classes of chosen tweets. The testing results show that the accuracy given by testing the 4000-tweets dataset reached 87%.

TABLE X.   DATASETS USED FOR TRAINING AND TESTING

Dataset	Positive	Negative	Total
Training	9666 (Sanders)	9666 (Sanders)	19,332
Testing	2000 (random)	2000 (random)	4,000

V. CONCLUSION AND FUTURE WORK

The social media becomes a reality in people's lives, enabling the growth of many online services. However, the companies maintain and assess the quality of their products or services by analyzing customers' satisfaction through social media platforms. The objective of this work is to propose a system that measures customer's satisfaction using sentiment analysis. Hence, the sentiment analysis is an important phase in the decision making process. We used the SVM as a classification algorithm beside the unigram as a feature extraction method and applied them to measure sentiment in Twitter data. The experimental result indicates that the unigram feature extraction method with SVM classification together bring high score reaches 87%. However, this percentage needs improvement either by using different dataset or different classification algorithm. As a future work, we can test other different classification algorithms and implement different feature extraction in addition to unigram. Moreover, we plan to specialize preprocessing and classification on medical or technology industries as they have definite glossary so the accuracy of the classification will be increased and become more focused. Finally, the algorithm will be applied on the other social media platforms such as Facebook, Instagram and Youtube.

REFERENCES

[1] M. Rouse, "Social Media Analytics," TechTarget, November 2012.

[2] F. K. Gohar, "Social Media Network Analytics," in Seven Layers of Social Media Analytics: Mining Business Insights from Social Media, 2015.

[3] L. Bing, "Sentiment Analysis and Subjectivity," in Handbook of Natural Language Processing, Illinois - Chicago, 2010, pp. 1-38.

[4] F. Weiguo and G. Michael, "The Power of Social Media Analytics," Communications of the ACM, pp. Vol. 57 Issue 6, PP. 74-81, 2014.

[5] O. Chong, S. Sheila and S. Almahmoud, "Social Media Analytics Framework: The Case Of Twitter And Super Bowl Ads," Journal of Information Technology Management, pp. 1-18, 2015.

[6] E. Younis, "Sentiment Analysis and Text Mining for Social Media Microblogs using Open Source Tools: An Empirical Study," International Journal of Computer Applications, pp. Vol. 112 , No. 5, PP. 44-48, February 2015.

[7] Z. Thabit, "Data Mining in Social Media," International Journal of Scientific and Engineering Research, vol. 6, no. 7, pp. Vol. 6, Issue 7,152-154, July 2015.

[8] P. Priyanka and M. Khushali, "A Review: Text Classification on Social Media Data," IOSR Journal of Computer Engineering, vol. 17, no. 1, 2015.

[9] K. Faris and K. Jugal, "Classifying Short Text in Social Media: Twitter as Case Study," International Journal of Computer Applications, February 2015.

[10] M. Isaac, "Mining Social Media For Predictive Analytics," Kampala, Uganda: School Of Computing and Engineering, Uganda Technology and Management University, 2015.

[11] toshiakit, "AnalyzeTwitter," 2014. [Online]. Available: https://github.com/toshiakit/AnalyzeTwitter. [Accessed 31 October 2016].

[12] J. Kunal, "Mining YouTube using Python and performing social media analysis," ALS ice bucket challenge, 2014.

[13] P. Bo and L. Lillian, "Opinion mining and sentiment analysis," Foundations and Trends in Information Retrieval, pp. Vol. 2, Nos. 1–2 (1–135), 2008.

[14] J. La, "Comparison of data analysis packages: R, Matlab, SciPy, Excel, SAS, SPSS, Stata," 23 February 2009. [Online]. Available: https://brenocon.com/blog/2009/02/comparison-of-data-analysis-packages-r-matlab-scipy-excel-sas-spss-stata/. [Accessed 31 October 2016].

[15] R. Matthew, Data Mining Facebook, Twitter, LinkedIn, Google+, GitHub, and More, O'reilly, 2014.

[16] K. Shamanth, M. Fred and L. Huan, Twitter Data Analytics, Springer, 2013.

[17] G. Chakraborty, M. Pagolu and S. Garla, "Text Mining and Analysis: Practical Methods, Examples, and Case Studies Using SAS. SAS Institute," in Text Mining and Analysis, 2013.

[18] B. Liu, Sentiment Analysis Opinion Mining, Morgan and Claypool Publishers, 2012.

[19] "Subjectivity Lexicon," MPQA , 2016. [Online]. Available: http://mpqa.cs.pitt.edu/lexicons/subj_lexicon/. [Accessed 30 october 2016].

[20] L. Naji, "Twitter Sentiment Analysis Training Corpus (Dataset)," 22 september 2012.

[21] K. P. Kishori and R. R. Deshmukh, "Twitter Sentiment Classification on Sanders Data using Hybrid approach," IOSR Journal of Computer Engineering , pp. Volume 17, Issue 4,PP. 118-123, 2015.

[22] T. AzureML , Microsoft, 2 September 2014. [Online]. Available: https://gallery.cortanaintelligence.com/Experiment/Binary-Classification-Twitter-sentiment-analysis-4. [Accessed 31 october 2016].

[23] B. Joel, J. Fredrik, J. Carl and W. Anders, "Emotion classification of social media posts for estimating people's reactions to communicated alert messages during crises," Security Informatics, vol. 3, no. 1, 28 august 2014.

[24] S. Hridoy, M. Ekram, M. Islam and R. M. Rahman, "Localized twitter opinion mining using sentiment analysis," Decision Analytics, vol. 2, no. 1, 22 october 2015.

[25] G. Pulkit and D. Sapan, Data Mining and Analysis on Twitter, 1st ed., ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE, 2011.

[26] H. Julian, "Build your own Twitter Archive and Analyzing Infrastructure with MongoDB, Java and R [Part 2] [Update] - ThinkToStart," 2013.

[27] J. Hillebrand, "Cluster Twitter Data with R and k-means - ThinkToStart," 13 may 2014. [Online]. Available: http://thinktostart.com/cluster-twitter-data-with-r-and-k-means/. [Accessed 31 october 2016].

[28] A. Bailey and C. Sufandha, "Twitter Analytics Using R Part 3: Compare Sentiments - www.credera.com," 4 June 2014. [Online]. Available: https://www.credera.com/blog/technology-insights/open-source-technology-insights/twitter-analytics-using-r-part-3-compare-sentiments/. [Accessed 31 october 2016].

[29] M. Bonzanini, "Mining Twitter Data with Python Part 1: Collecting Data," Kdnuggets.com, 2015. [Online]. Available: http://www.kdnuggets.com/2016/06/mining-twitter-data-python-part-1.html.. [Accessed 31 october 2016].

[30] L. Shure, "Analyzing Twitter with MATLAB » Loren on the Art of MATLAB," Blogs.mathworks.com, 2014. [Online]. Available: http://blogs.mathworks.com/loren/2014/06/04/analyzing-twitter-with-matlab/. [Accessed 31 october 2016].

[31] M. Zaikin, "Common Architectures: Explain the advantages and disadvantages of multi-tier architectures," in Java Design: Objects, UML, and Process, Robert C. Martin Series, 2007.

[32] A. Agarwal, "Sentiment Analysis of Twitter Data," in In: Proc. ACL 2011 Workshop on Languages in Social Media, 2010 .

[33] B. Carver, "5 Reasons to Invest in Dedicated Social Media Analytics Tools," 2016. [Online]. Available: https://www.linkedin.com/pulse/5-reasons-invest-dedicated-social-media-analytics-tools-bob-carver. [Accessed 31 October 2016].