

Opinions Mining in Facebook

Dr. Mohammed Hajjouz
Albaath University, Homs, Syria

Abstract

Because of the tremendous development in all areas of scientific, economic, social and others, the need emerged to find unconventional ways in which we deal with text data that have become very large sizes these days. Therefore, we must find new ways to derive knowledge and hidden information within these huge amounts of data, and these techniques in exploration data that we used in this study are classification techniques. We fetch the data from Facebook social networking site, and then we have worked on cleaning and preparing the text to texts classification process. These texts are contained a lot of noise and information is useful for the opinions analyzing process, such as advertisements, links, e-mail addresses, and the presence of many words that do not effect on the general orientation of the text. After we get all the posts from the Facebook page and its related comments, our goal is to know the Percentage of positives and negatives for this post. We applied Naïve Bayes algorithm in classification, we had the appropriate training, and after passing Posts and comments data (opinions), we got good results on the ratio of positives and negatives of the.

Keywords: Opinions mining, Sentiment classification, Challenge, Access token

1. Introduction

The emergence of new specialized data collection technologies lead to the data accumulation on a large scale in various areas. In addition, the conventional query methods for data are not sufficient to extract useful information from large databases. The human capacity has become limited to store this vast amount of information without the use of techniques, which help him to understand the data and extract knowledge of them. As a result, the data mining appeared because of the urgent need to deal with large volumes of data with various exploration techniques including classification technique used in the completion of this research. The classification technique classified in the field of supervised learning based on Prior Information to predict the future, so classification techniques extract or derive models that define categories or classes of objects depending on the features and attributes by these classes and which been identified in the stage of learning or training.

2. The aim of the research:

This research aims to analyze the opinions of commentators on the social networking site Facebook in order to evaluate any opinion or commodity like (University, Mobile, etc.) using classification technology and machine learning approach, which is one of the core functions of data mining used in a broader process to explore knowledge. This research helps the user to access the evaluation of other users through Posts and commits on the social networking site for any opinion of the immediate and automatic it is then the process of analyzing and evaluation the opinions using the appropriate algorithms for this purpose.

3. Classification of data mining:

There are several categories in the data mining:

- classification
- Prediction
- Summarization
- Clustering
- Association Rule
- Anomaly detection

Our research use classification property that is classified within the field of supervised learning based on advance information in order to predict future information. The classification process helped to extraction or devising models that define categories or classes of objects depending on the features and attributes enjoyed by these objects and that have been learned in the learning phase or training.

The classification process can complete on the old statistical methods (such as regression analysis and discriminatory) It is best to use a relatively new methods such as (association rules, the conclusion based case and neural networks). Examples, which use classification, work exploring knowledge applications that include the classification of financial market trends^[1].

4. Data mining algorithms:

There are many algorithms used data mining process, can choosing the appropriate algorithm depending on the nature of the problem. The most important algorithms used in data mining:

- Nearest Neighbor.
- Decision Trees.
- Naïve Bayesian.
- Clustering.
- Association Rules.
- Neural Network.
- Time Series.
- Support Vector Machines.

Used in our research this Naïve Bayesian algorithm, which is one of the most important algorithms in the field of mining data. This algorithm is easy to implementation and the ability to handle huge data and this algorithm based on the theory of the world Bayes ^[1].

5. Opinions Mining:

Most human activities rely on the opinions formed for the persons, because it has a big impact on our beliefs and our behavior and our perception of reality, and the options that we selected. Therefore, when we have a problem or want to take a specific decision to ask others to give their opinions, and apply this truth is not only the people but also the organizations and institutions. In the past, few years have changed the way people express their opinions and feelings radically thanks to social networks and virtual communities and other social media appearance on the internet.

The knowledge of the discovery of a huge amount of unstructured data on the web has become one of the important challenges nowadays due to its importance in different areas of life. For example, can be a major factor for marketers who want to create an image or identity in the minds of their customers for their products and commercial grades. Involves the automatic analysis of opinions in fact a deep understanding of natural language text by hardware and this soon became the target of the verification.

Analyzing the opinions and all the concepts related to it, such as feelings, evaluations and attitudes are all emotions materials offered for study by analyzing opinions technique. With the fast growth of the communication and social media such as forums and social networks. We have for the first time in human history has become such a huge amount of information and data recorded and saved in digital form. Since 2000 became the analysis of the opinions evolve into one of the most affected areas in the field of natural language processing and data mining ^[2].

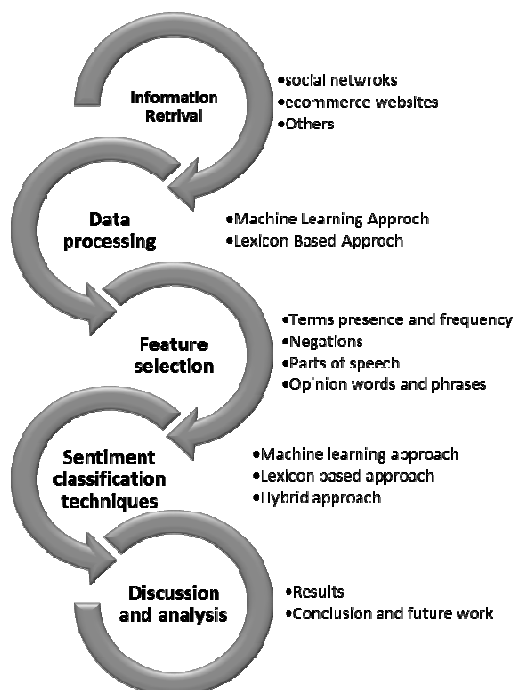


Figure 1. The global diagram for opinions mining.

5.1 Sentiment Analysis and Opinion Mining:

Sentiment Analysis (SA) or Opinion Mining (OM) is the computational study of people's opinions, attitudes and emotions toward an entity. The entity can represent individuals, events or topics. These topics are most likely to be covered by reviews. The two expressions SA or OM are interchangeable. They express a mutual meaning.

Opinion Mining extracts and analyzes people’s opinion about an entity while Sentiment Analysis identifies the sentiment expressed in a text then analyzes it. Therefore, the target of SA is to find opinions, identify the sentiments they express, and then classify their polarity.

5.2 Classification levels in Sentiment Analysis:

There are three main classification:

- 1- Document-level: aims to classify an opinion document as expressing a positive or negative opinion or sentiment. It considers the whole document a basic information unit.
- 2- Sentence-level: aims to classify sentiment expressed in each sentence. The first step is to identify whether the sentence is subjective or objective. If the sentence is subjective, Sentence-level SA will determine whether the sentence expresses positive or negative opinions.
- 3- Aspect-level: aims to classify the sentiment with respect to the specific aspects of entities. The first step is to identify the entities and their aspects. The opinion holders can give different opinions for different aspects of the same entity.

5.3 Feature selection in sentiment classification:

Sentiment Analysis task is considered a sentiment classification problem. The first step in this problem is to extract and select text features ^[4].

- Terms presence and frequency: These features are individual words or word n-grams and their frequency counts. It gives either the words binary weighting (zero if the word appears or one if otherwise) or uses term frequency weights to indicate the relative importance of features Parts of speech (POS).
- Finding adjectives, as they are important indicators of opinions.
- Opinion words and phrases: these are words commonly used to express opinions including good or bad, like or hate. On the other hand, some phrases express opinions without using opinion words.
- Negations: the appearance of negative words may change the opinion orientation like not good is equivalent to bad.

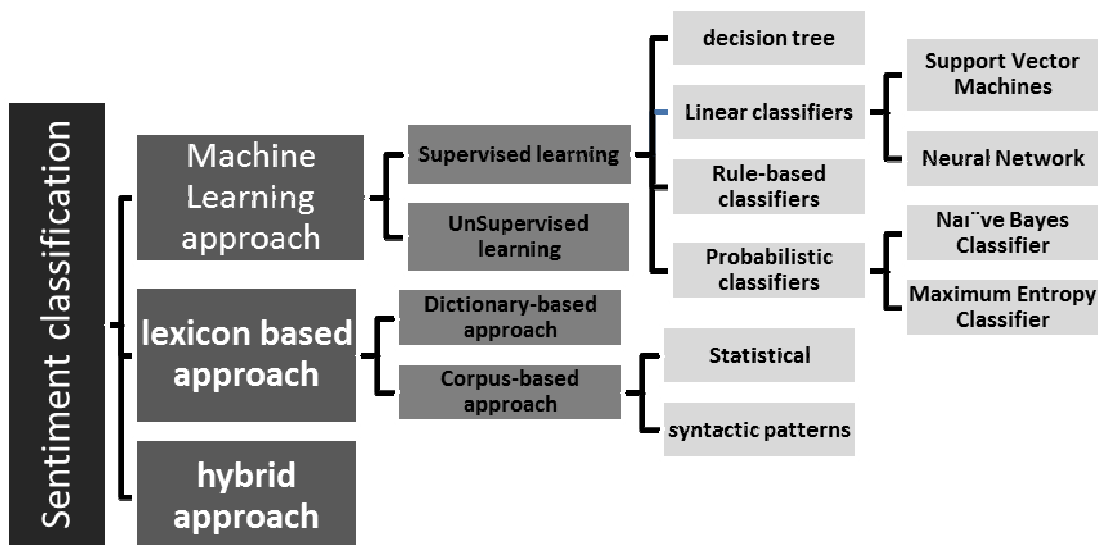


Figure 2. Sentiment classification techniques.

Feature Selection methods are divided in tow based methods ^[4]:

- Lexicon-based approaches, need human annotation, usually working with a small group of words that form the prelude to the discovery of synonyms process in order to get a larger dictionaries starts, but this method is facing a lot of trouble.
- Statistical approaches, which are automatic methods that are more frequently used.

5.4 Sentiment classification techniques:

There are three Sentiment classification techniques:

5.4.1 Machine learning approach:

Machine learning approach relies on the famous Machine learning algorithms to solve the Sentiment Analysis as a regular text classification problem that makes use of syntactic and/or linguistic features.

Text Classification Problem Definition: We have a set of training records $D = \{X_1, X_2, \dots, X_n\}$ where

each record is labeled to a class. The classification model is related to the features in the underlying record to one of the class labels.

Then for a given instance of unknown class, the model is used to predict a class label for it. The hard classification problem is when only one label is assigned to an instance. The soft classification problem is when a probabilistic value of labels is assigned to an instance.

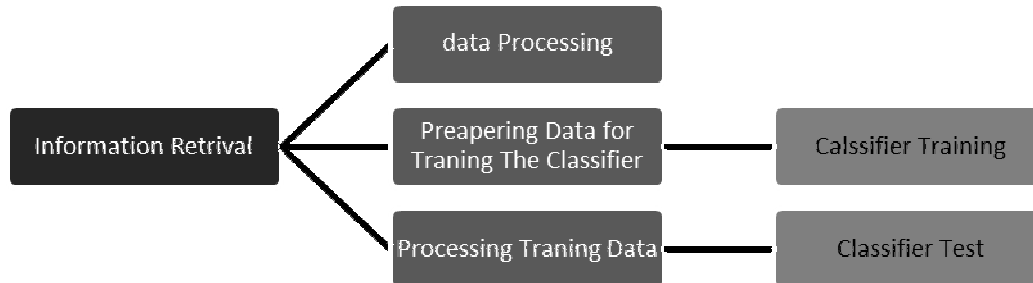


Figure 3. Machine learning approach

Of the most important Machine learning classifiers under supervised:

1. Probabilistic classifiers: Probabilistic classifiers use mixture models for classification. The mixture model assumes that each class is a component of the mixture. Each mixture component is a generative model that provides the probability of sampling a particular term for that component. These kinds of classifiers also called generative classifiers.
 - 1.1 Naïve Bayes Classifier (NB). The Naïve Bayes classifier is the simplest and most commonly used classifier. Naïve Bayes classification model computes the posterior probability of a class, based on the distribution of the words in the document. The model works with the BOWs feature extraction, which ignores the position of the word in the document. It uses Bayes Theorem to predict the probability that a given feature set belongs to a particular label.
 - 1.2 Maximum Entropy Classifier (ME). The maximum classifier (known as a conditional exponential classifier) converts labeled feature sets to vectors using encoding. This encoded vector then used to calculate weights for each feature that can be combined to determine the most likely label for a feature set. This classifier is parameterized by a set of $X\{\text{weights}\}$, which is used to combine the joint features that are generated from a feature-set by an $X\{\text{encoding}\}$. In particular, the encoding maps each $C\{\{\text{feature set, label}\}\}$ pair to a vector.
 2. Linear classifiers:
 - 2.1 Support Vector Machines Classifiers (SVM). The main principle of SVMs is to determine linear separators in the search space, which can best separate the different classes.
 - 2.2 Neural Network (NN). Neural Network consists of many neurons where the neuron is its basic unit. The vector denotes the inputs to the neurons over line X_i that is the word frequencies in the document.
 3. Decision tree classifiers. Decision tree classifier provides a hierarchical decomposition of the training data space in which a condition on the attribute value is used to divide the data
 4. Rule-based classifiers: In rule-based classifiers, the data space is modeled with a set of rules. The left hand side represents a condition on the feature set expressed in disjunctive normal form while the right hand side is the class label. The conditions are on the term presence. Term absence is rarely used because it is not informative in sparse data.
- 5.4.2 Lexicon-based approach:
- Opinion words are employed in many sentiment classification tasks. Positive opinion words are used to express some desired states, while negative opinion words are used to express some undesired states. There are also opinion phrases and idioms, which together are called opinion lexicon ^[5].
- There are three main approaches:
1. Dictionary-based approach: presented the main strategy of the dictionary-based approach. A small set of opinion words is collected manually with known orientations. Then, this set is grown by searching in the well-known corpora.
 2. Corpus-based approach: The Corpus-based approach helps to solve the problem of finding opinion words with context specific orientations. Its methods depend on syntactic patterns or patterns that occur together along with a see list of opinion words to find other opinion words in a large corpus.
 3. Statistical approach: Depends on statistical methods to identify patterns and a group of words Opinion
- 5.4.3 Hybrid approach: mixed from the last two types.

5.5 Data processing:

Data processing involves many steps as following:

5.5.1 Operations on text: It includes retrieved the data from the Facebook and cleaning it, and processing of text process for the classification process. Texts retrieved contain a lot of noise and useful information for the process of the opinions mining, such as advertisements and links and e-mail addresses and the presence of many words that do not affect the general orientation of the text [6].

The total process of data preparing include many steps:

- Tokenization
- Remove HTML Tags
- Remove Hash Tags
- Remove Numbers
- Remove Links and Emails
- Remove dates, month names and day names

5.5.2 Part-of-Speech Tagging: The process of assigning a part-of-speech to each word in a sentence.

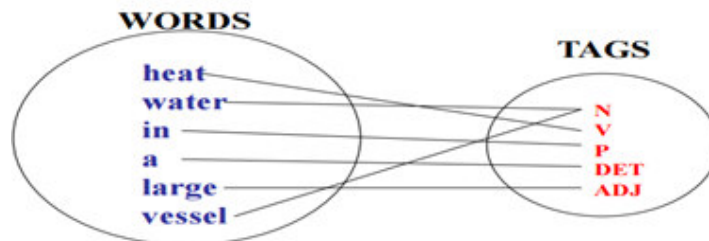


Figure 4. Classification of the word grammar.

5.5.3 Depending on the model that we will be built by extracting all the Verbs and Adjective and Adverb that belong to a group of speech set out in the table.

JJ	Adjective
JJR	Adjective, comparative
JJS	Adjective, superlative
VB	Verb, base form
VBD	Verb, past tense
VBN	Verb, past participle
VBP	Verb, non-3rd person singular present
VBZ	Verb, 3rd person singular present
RB	Adverb
RBR	Adverb, comparative
RBS	Adverb, superlative
VBG	Verb, gerund or present participle

For do this, we relying on the next steps:

1- Adjectives :

Basic (JJ)	red, tall
Comparative (JJR)	redder, taller
Superlative (JJS)	reddest, tallest

2- Verbs :

Base, infinitive (VB)	Eat
Past tense (VBD)	Ate
Gerund (VBG)	Eating
Past participle (VBN):	Eaten
3 rd person singular present tense: (VBZ)	Eats

3- Adverb:

Basic (RB):	Quickly
Comparative (RBR):	Quicker
Superlative (RBS):	Quickest

We are relying on these words that extracted as features then classification process will be apply because these words have a very large and crucial role in the process of expression.

The process of identifying parts of speech through the application of Tokenization process. That make segmenting sentence to the list of individual words and then apply the process of identifying parts of speech on these words to define formulas grammatical corresponding and the output of the process of parts of speech is the

input of the classifier, which will perform text classification. The following example shows the process of identifying parts of speech:

The movie was great I enjoyed it	
Movie	NN
Great	JJ
Enjoyed	VBD

Then the output of this stage is the input for the next stage.

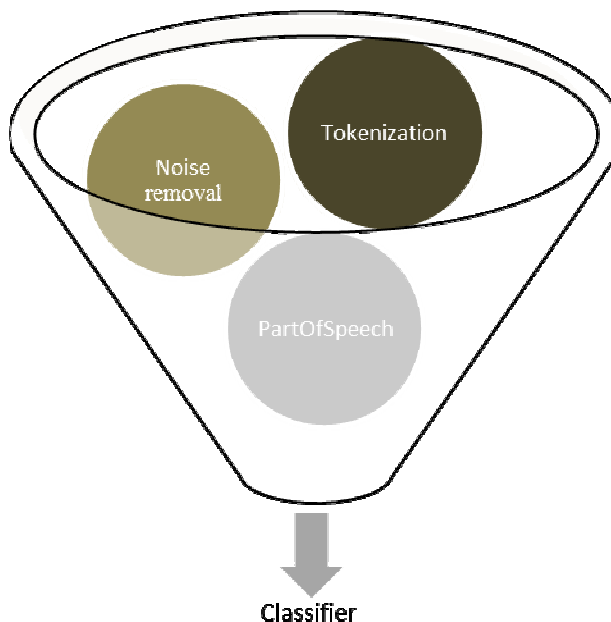


Figure 5. Classifier work mechanism.

6. Naïve Bayes Algorithm

The Bayesian Classification represents a supervised learning method as well as a statistical method for classification. Assumes an underlying probabilistic model and it allows us to capture uncertainty about the model in a principled way by determining probabilities of the outcomes. It can solve diagnostic and supervised problems. This Classification is named after Thomas Bayes (1702-1761) who proposed the Bayes Theorem.

Bayesian classification provides practical learning algorithms and prior knowledge and observed data can be combined. Bayesian Classification provides a useful perspective for understanding and evaluating many learning algorithms. It calculates explicit probabilities for hypothesis and it is robust to noise in input data^[7].

6.1 Naïve Bayes classifier:

A Naïve Bayes classifier is a simple probabilistic classifier based on applying Bayes theorem (from Bayesian statistics) with strong (naïve) independence assumptions. A more descriptive term for the underlying probability model would be "independent feature model". In simple terms, a Naïve Bayes classifier assumes that the presence (or absence) of a particular feature of a class is unrelated to the presence (or absence) of any other feature^[7].

Depending on the precise nature of the probability model, Naïve Bayes classifiers can be trained very efficiently in a supervised learning setting. In many practical applications, parameter estimation for Naïve Bayes models uses the method of maximum likelihood; in other words, one can work with the Naïve Bayes model without believing in Bayesian probability or using any Bayesian methods.

An advantage of the Naïve Bayes classifier is that it only requires a small amount of training data to estimate the parameters (means and variances of the variables) necessary for classification. Because independent variables are assumed only the variances of the variables for each class need to be determined and not the entire covariance matrix.

Bayes' Rule Applied to Documents and Classes:

- For a document d and a class C :

$$P(C|d) = \frac{P(d|C)P(C)}{P(d)} = P(d|C)P(C)$$

$P(C)$: The previous probability of class C .

$P(d|C)$: The probability of document d for class C .

$P(d)$: The previous probability of document d for class C .

• Naïve Bayes Independence Assumptions:

1. Bag of words assumption: Assume position does not matter.

2. Conditional Independence Assume the feature probabilities $P(x_i|c_j)$ are independent given the class c .

$$P(x_1, x_2, \dots, x_n | C) = P(x_1 | C) * P(x_2 | C) * \dots * P(x_n | C)$$

3. If we have d document that contains words (W_i) and two classes C_1, C_2 we can calculate the probability of the class for the document d by the relationship:

$$P(W_1, W_2, \dots, W_n | C_1) = P(C_1)P(W_1 | C_1) * P(W_2 | C_1) * \dots * P(W_n | C_1)$$

$$P(W_1, W_2, \dots, W_n | C_2) = P(C_2)P(W_1 | C_2) * P(W_2 | C_2) * \dots * P(W_n | C_2)$$

4. If the probability of class C_2 greater than the probability of C_1 , by then the class C_2 is belongs to the document d .

The possibilities calculating process accordance with the following:

First attempt: maximum likelihood estimates

1. Calculate the previous probability for class C :

$$P(C_j) = \frac{D_{occount}(C = c_j)}{N_{doc}}$$

$D_{occount}(C = c_j)$: The number of documents in class C .

2. The possibility of the word in the class C :

$$P(W_i | C_j) = \frac{count(W_i, c_j)}{\sum_{w \in V} count(w, c_j)}$$

$count(W_i, c_j)$: The frequency of the word (W_i) in class C .

$\sum_{w \in V} count(w, c_j)$: The total of words in a class C .

Problem with maximum likelihood:

What if we have seen no training documents with the word fantastic and classified in the topic positive. For instance, if we have, "great" word in the test text and does not appear in training documents. Therefore, we have:

$$P(w | pos) = \frac{count(great, pos)}{\sum_{w \in V} count(w, pos)} = 0$$

Thus, we are in front of a big problem so that the probability of the class to which it belongs text laboratory equal to (zero) thus do not get any result and the solution in such cases is the process of adding (1) to the denominator and numerator as following:

$$P(W_i | C_j) = \frac{count(W_i, c_j) + 1}{\sum_{w \in V} count(w, c_j) + 1} = \frac{count(W_i, c_j) + 1}{\sum_{w \in V} count(w, c_j) + |V|}$$

Below a practical example illustrates the work mechanism Naïve Bayes classifier:

If we have, two class (POS) and (NEG). The class (POS), contains documents that contain words that express positive opinion and the class (NEG), contains documents that contain words that express a negative opinion as following:

	Documents	Words	Class
Training	1	good nice good	POS
	2	good good cool	POS
	3	good fantastic	POS
	4	Bad terrible good	NEG
Test	D5	Good good good bad terrible	???

$$1- P(POS) = \frac{Doccount(POS)}{N_{doc}} = \frac{3}{4} \quad P(NEG) = \frac{Doccount(NEG)}{N_{doc}} = \frac{1}{4}$$

$$2- P(good | POS) = \frac{count(good, POS) + 1}{\sum_{w \in V} count(w, POS) + |V|} = (5 + 1) / (8 + 6) = \frac{3}{7}$$

$$P(bad | POS) = \frac{count(bad, POS) + 1}{\sum_{w \in V} count(w, POS) + |V|} = (0 + 1) / (8 + 6) = \frac{1}{14}$$

$$P(terrible | POS) = \frac{count(terrible, POS) + 1}{\sum_{w \in V} count(w, POS) + |V|} = (0 + 1) / (8 + 6) = \frac{1}{14}$$

$$P(good | NEG) = \frac{count(good, NEG) + 1}{\sum_{w \in V} count(w, NEG) + |V|} = (1 + 1) / (3 + 6) = \frac{2}{9}$$

$$P(bad | NEG) = \frac{count(bad, NEG) + 1}{\sum_{w \in V} count(w, NEG) + |V|} = (1 + 1) / (3 + 6) = \frac{2}{9}$$

$$P(terrible | NEG) = \frac{count(terrible, NEG) + 1}{\sum_{w \in V} count(w, NEG) + |V|} = (1 + 1) / (3 + 6) = \frac{2}{9}$$

$$3- P(POS | D5) = \frac{3}{4} * \frac{3}{7} * \frac{3}{7} * \frac{3}{7} * \frac{1}{4} * \frac{1}{4} = 0.0003$$

$$4- P(NEG | D5) = \frac{1}{4} * \frac{2}{9} * \frac{2}{9} * \frac{2}{9} * \frac{2}{9} * \frac{2}{9} = 0.0001$$

5- If POS > NEG → POS

6- IF POS < NEG → NEG

6-2 Advantages/Disadvantages of Naïve Bayes [7]:

6.2 Advantages:

- Fast to train (single scan). Fast to classify
- Not sensitive to irrelevant features
- Handles real and discrete data
- Handles streaming data well
- Disadvantages:
- Assumes independence of features

6-3 Precision, Recall, F-measure:

	Positive	Negative
	Correct	Not Correct
Selected(POS)	tp	fp
not selected(NEG)	fn	tn

Precision: Percentage of selected items that are correct:

$$Precision = \frac{tp}{tp + fp}$$

Recall: percentage of correct items that are selected:

$$Recall = \frac{tp}{tp + fn}$$

F- measure:

$$F = \frac{2 * precision * recall}{precision + recall}$$

7. The Facebook

The Facebook website of the most popular social networks at the moment and it change the traditional ways between people to express their views and communication because of the ease of use and the services provided by it.

The social network accounted for the acceptance and response of many people, especially young people all over the world. Even it does not exceed the limits of personal blog at the beginning of its inception in February (2004), at Harvard University in the United States, by a student named *Mark Zuckerberg*, and his blog confined in the scope of at the university and limits of friends Zuckerberg,

Then the Facebook crossed the boundaries of the university fame spread in the different schools of the United States of America.

The developers of Facebook in (2007) added new capabilities to this site and which provide an opportunity for developers than the popularity of Facebook site, so exceeding the limits of the United States to all countries of the world.

The number of visitors to this site on the first of July (2010), half a billion people, visit constantly and share among themselves files, photos, videos, and comment on what is published in the pages of opinions and ideas and a variety of topics and new. In addition, the active participation and are often in conversations and chats.

The Facebook site occupies of fame and popularity third place after (Google & Microsoft), and the number of visitors more than a billion users.

7.1 A mechanism to access the data stored in the Facebook site:

We will show how we have access to the data available on the Facebook site, which consists of posts within a certain page and comments related.

- (Facebook Graph API) Considered as The basic method for data acquisition and entry to the site, through the use of query or add new posts or advertising management or upload images and other tasks that we need, which is comprised of the following components:
- Nods: the things that are on the site. Such as user or images or pages or comments.
- Edges: They express the existing connections between the nodes.
- Fields: contains information that reflect the nods, such as the page name.

We can access the Facebook Graph API and do various available tasks; by creating, an application on Facebook uses the (access token).

Access token is a vague series of characters containing information about the time period allowed for the usage of (access token) and also information the application was created the (access token), it determines the user's identity or the application or page in order to request queries from the Facebook Graph API [8].

7.2 Import data from Facebook:

The goal is to retrieve the required data from the Facebook site, so we have been building this section to fetch data from Facebook site using the Facebook SDK for .NET. Which helps developers to build different applications integrated with the Facebook site.

- Establishing application on <https://developers.facebook.com> site characterized this process as easy and after the completion of creating the application we can generate (access token) that allow us to process communication and authentication between the application and the Facebook Graph API.
- We build `DataFromFacebook.cs` class and define within it variable by type a `FacebookClient` using Facebook library located within the Facebook SDK for .NET in order to obtain (access token). After obtaining the (access token) now it possible to send requests to the Facebook Graph API and fetch the required data:

```
class DataFromFacebook
{
// Facebook Library
FacebookClient FB;
Public string Get_Access_Token()
{
try
{
```

```
var fb = new FacebookClient(); //Login without token access
    dynamic result = fb.Get("oauth/access_token", new
    {
        client_id = "369673403185791",
        client_secret = "5ff96468780678b1d6b1c23cb82e682f",
        grant_type = "client_credentials"
    });

    string FaceBook_A_Token = result.access_token;
    return FaceBook_A_Token;
}
catch (Exception)
{
    return null;
}
}
public dynamic Get_Pages_Posts(string accessToken, string PageName)
{
    Try
    {
        //brink the posts page
        FB = new FacebookClient(accessToken);
        dynamic Posts = FB.Get(PageName + "/posts?fields=message");
        return Posts;
    }
    catch (Exception)
    {
        return null;
    }
}

public dynamic Get_Comments(string accessToken, string PostId)
{
    try
    {
        //brink the comments
        FB = new FacebookClient(accessToken);
        dynamic Commants = FB.Get(PostId + "?fields=comments");
        return Commants;
    }
    catch (Exception)
    {
        return null;
    }
}
```

8. The mechanism work of research:

Below we explain the steps involved in completing the research:

8.1 Usecase:

The figure (6) shows Usecase scheme, where the user connects to Facebook location and bring the posts and comments specialized in each post and then analyzed.

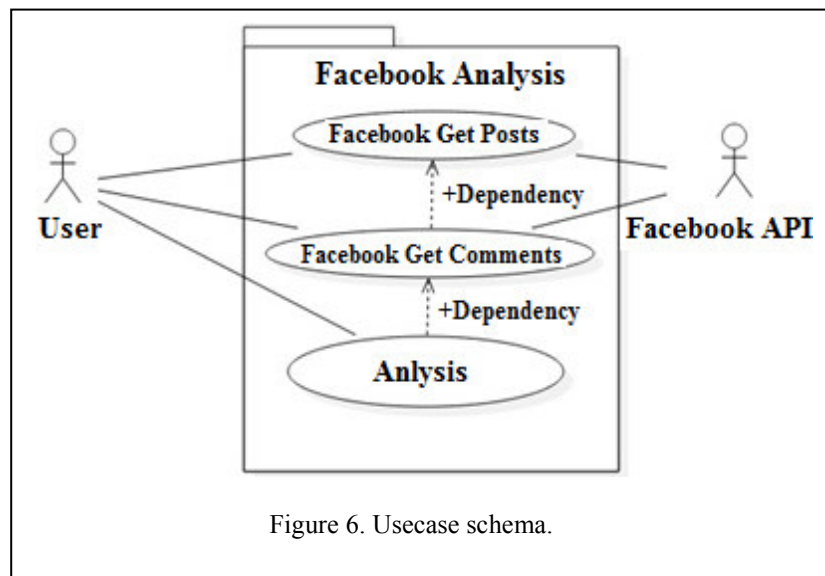


Figure 6. Usecase schema.

8.2 Access to publications from Facebook:

Facebook UC01 = Get Posts			
The main scenario	Get Posts		
Trigger			
Actor	User Facebook Api		
Related use cases	Internet connection – effective ACCESS Token		
Preconditions	Internet connection - ACCESS Token		
Chain of events	User	System	Facebook Api
	<ul style="list-style-type: none"> User enter a Facebook page link. Press Get Posts 	<ul style="list-style-type: none"> The system opens a channel of communication with the Facebook Api Access system sends a special Access token related to Facebook application that we have created on the site 	<ul style="list-style-type: none"> The Facebook Api Access Token verifies the validity of the sender restore the required data
An alternative scenario	Wrong or expired Access Token. No active Internet connection. No comments on the publication selected by the user		

8.3 Get comments from posts in Facebook:

Facebook UC02 = Get Comments			
The main scenario	User review posts and press Fetch Comment. The system displays associated comments with the post chosen by the user		
Trigger	Press Fetch Comment		
Actor	User – Facebook Api		
Related use cases	Facebook Get Posts		
Preconditions	Internet connection - ACCESS Token		
Chain of events	user	system	Facebook Api
	<ul style="list-style-type: none"> Reviewed the publications displayed to a user and select of them press Fetch Comment 	<ul style="list-style-type: none"> The system opening a channel of communication with the Facebook Api the system send the Access token related to Facebook application that created The system sends the ID number publication chosen by the user 	<ul style="list-style-type: none"> The Facebook Api verifies the Access Token the validity The Facebook Api validates the ID publication sender Return the required data.
An alternative scenario	Wrong or expired Access Token. No active Internet connection. No comments on the publication selected by the user		

8.4 Analysis of the comments that have been obtained from the Facebook pages:

UC03 = Facebook Analysis	
The main scenario	The user presses the Analysis button
Trigger	The user presses the Analysis button
Actor	User
Related use cases	Facebook Get Posts. Facebook Get Comments.
Preconditions	The existence of comments on the posts
Chain of events	User
	<p>The user presses the Analysis button</p>
	System
	<p>The system stores the comments displayed in the program interface.</p> <p>POS</p> <p>Deletion of the words that do not reflect the opinion.</p> <p>Delete numbers and links.</p> <p>Text words are classified according to linguistic classifications in the English language rules are selected classifications required.</p> <p>The output text from the previous step pass on the special laws of algorithm.</p> <p>Connect to the database and retrieve information on every word of the text words (rating - the number of times the roses - the number of iterations).</p> <p>Calculate Positive and negative likelihood each comment.</p> <p>The algorithm returns the number of positive comments and the number of negative comments.</p> <p>Previous numbers are converted to percentages show to the user.</p>
An alternative scenario	Wrong or expired Access Token No active Internet connection. No comments on the publication selected by the user. Not all existing comments reflect the opinion of commentators. All comments are written in a language other than English.

8.5 Schema Facebook Activity:

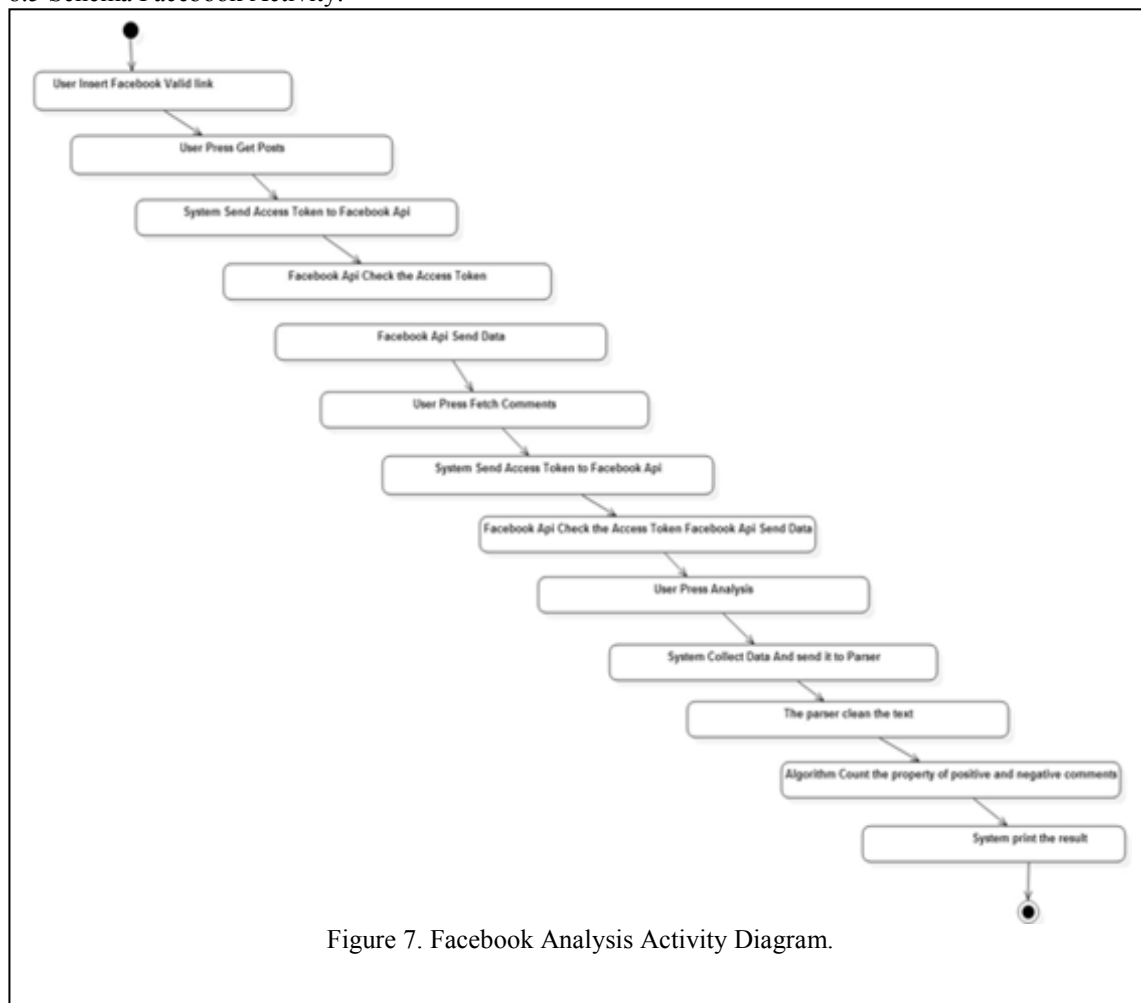


Figure 7. Facebook Analysis Activity Diagram.

9. Results

Using the above ideas were built software application works to bring all posts and the comments of any English Facebook page, then analyze these opinions.

After pass, training base to algorithm the application has been experimenting for a certain period and the results were somewhat acceptable with some errors. A review of the application has been done to correct these errors.

Then we were overcome these errors and tested it again, we obtain almost the proportion of correct results 97% of the test sample. These results are considered excellent, especially that we are dealing with social networking sites where users write their opinions using colloquial English that contain a lot of symbols and expressions that are difficult for an analyst analysis, the application gives the percentage of positive and negative opinions.

We Test the application on the guardian Facebook page, which have the following URL:

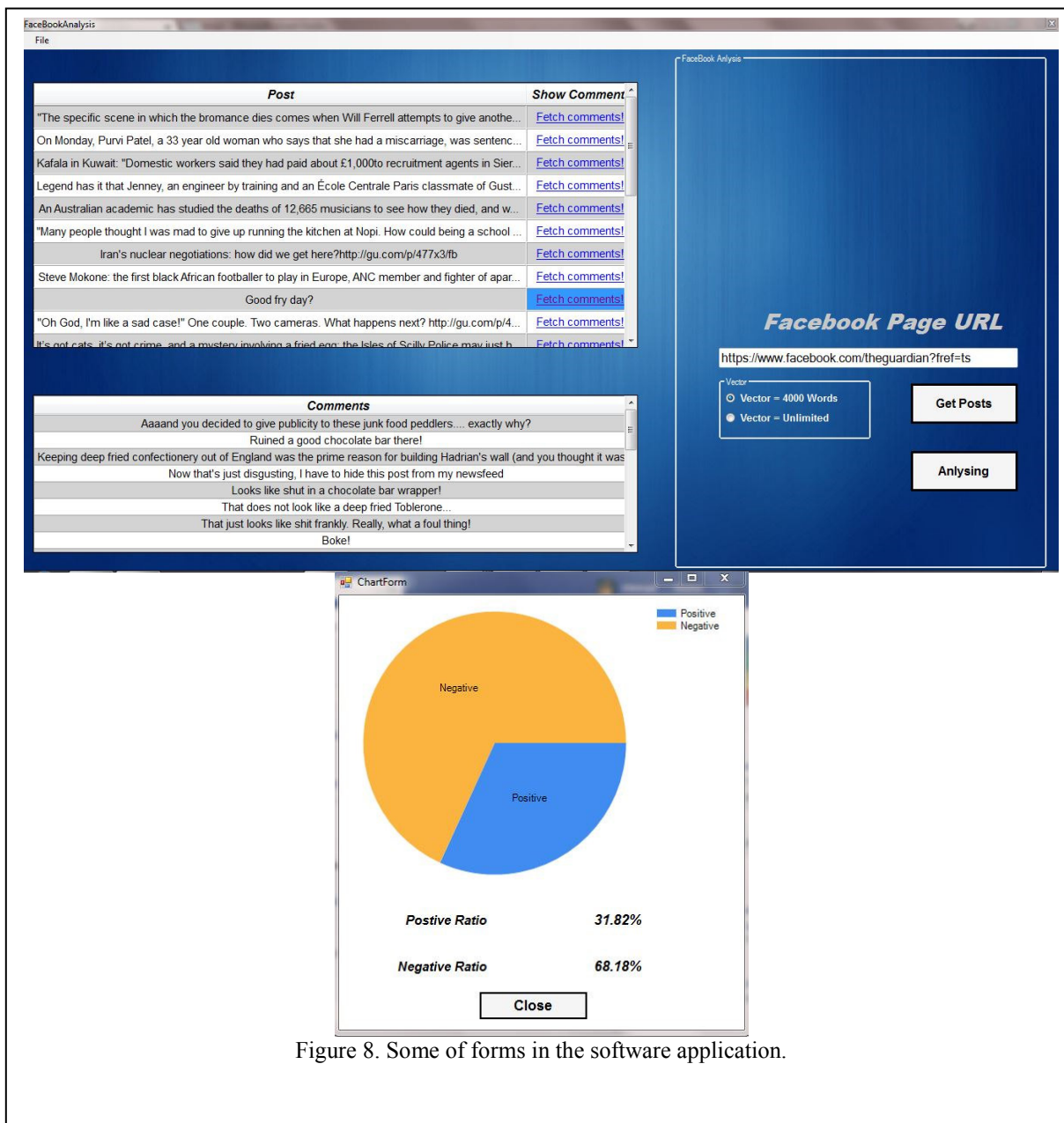
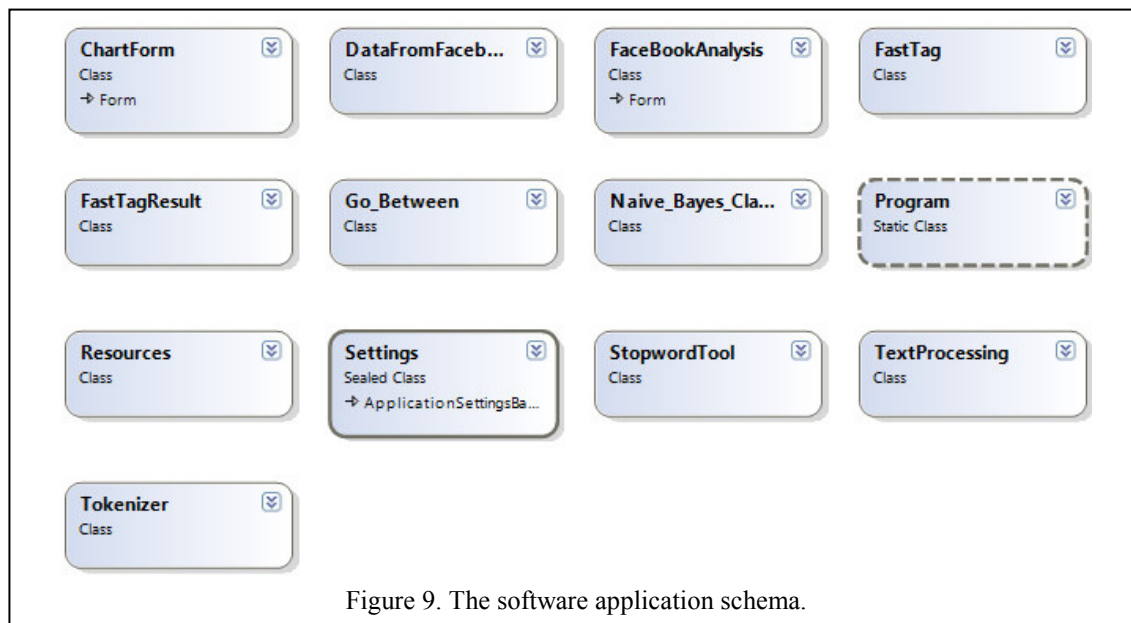


Figure 8. Some of forms in the software application.

("https://www.facebook.com/theguardian?ref=ts&fref=ts")



The Figure (10) shows the software application schema with its classes and models:

10. Conclusion:

Through this search, were able to analyze the opinions in Facebook and we have reached good results using Bayes algorithm. However, we hope that achieved the following objectives:

1. Application develop in order to analyze the opinions of most social networking sites such as Twitter, YouTube, Google Plus and Instagram.
2. Classification Arabic-language texts because it's our language and mother language of nearly 250 million people, and about 10 million people use the Internet, and this large number leads us to think about the development of the application to be able to deal with the opinions written in Arabic.
3. The current classification according with two models negative or positive. However, there are many opinions that contains contradictory feelings. It is possible that the writer expresses his admiration for a product, but is not impressed with one of its features, so must be developed this application in line with these cases.
4. The current application to classify the opinions of commentators on the social networking sites only. It is best to develop the application deals with text documents.
5. Its batter to use professional text parser like Stanford parser.

Note 1. You can download the application and try it through the following link:

http://www.4shared.com/rar/OBkcIRwEce/Opinions_Analysis_Setup.html

References:

- [1] Data Mining Concepts and Techniques Second Edition Jiawei Han and Micheline Kamber.
- [2] H. Tang, S. Tan, X. Cheng, A survey on sentiment detection of reviews, Expert Systems with Applications 36 (7) (2009).
- [3] Wilson T, Wiebe J, Hoffman P. Recognizing contextual polarity in phrase-level sentiment analysis.
- [4] Aggarwal Charu C, Zhai Cheng Xiang. Mining Text Data.
- [5] Diana Maynard, Adam Funk. Automatic detection of political opinions in tweets. In: Proceedings of the eighth international conference on the semantic web, ESWC'11; 2011. p. 88–99.
- [6] B. Pang, L. Lee, S. Vaithyanathan, Thumbs up? Sentiment classification using machine-learning techniques.
- [7] Naive Bayes classifiers Kevin P. Murphy 2006.
- [8] <https://developers.facebook.com/docs>.