

## An Application of Emotion Detection in Sentiment Analysis on Movie Reviews

Ria Ambrocio Sagum, MCS<sup>a,b</sup>

<sup>a,b</sup>College of Computer and Information Sciences, Research Management Office, Polytechnic University of the Philippines, Sta. Mesa, Manila

**Article History:** Received: 10 November 2020; Revised 12 January 2021 Accepted: 27 January 2021; Published online: 5 April 2021

---

**Abstract:** The research focus on the issue of accuracy for sentiment analysis. The researcher experimented on emotion detection result to be used in sentiment analysis. The emotions that were included in this research are happiness, sadness, anger, and fear. Once emotion was detected the system will then use it to know the sentiment of the person on a particular movie. This paper aims to measure the accuracy in sentiment analysis enhanced by emotion detection and to know whether emotion detection plays a key role in reading sentiment analysis.

**Keywords:** Emotion Detection, Language Processing, Sentiment Analysis on Movie Reviews, Emotion Detection on Movie Reviews

---

### 1. Introduction

To date the use of sentiment analysis in social media monitoring became the latest trend, because of this people gain an overview of the public opinion in a particular topic. The ability to extract insights from social data is common and is being adopted by organizations.

In the movie industry, some of the moviegoers check first the review of a particular movie before watching it in the cinema. A good review can sometimes be considered as a good box-office success. But one cannot rely on a person review alone, one needs to consider that different person has different insight based on their lifestyle and cultural belief. It is said that a movie can create change cultural climate of the viewer. A study of Persson [1], exemplifies that a cinema movie may introduce a new theme, lifestyle, fashion style, and or different conventions that will change the way the critics, authors, and audience understand its literature. As how cinema affect persons lifestyle and beliefs that give definition to human society, its semiotics and articulation differs. With this we can say that a movie can be seen by people differently based on their culture and lifestyle. Given this still sentiments of people doing the review is important and must be accurate to save time and money for some people. This time people do not want to expect something that cannot be given by the movie. A good sentiment analysis can help these people by simply looking on the sentiments result. For some reviews you can see it thru the stars, but sometimes these are just numerical. Sentiment analysis as application of Natural Language Processing (NLP) does not compute numerical feedback to give the sentiments rather use text analysis of the sentences based on the reviewers post or feedbacks.

There are existing sentiment analysis that are being used for reviews but researchers are still trying to find different ways to be able to accurately analyze a sentiment for a review. Different variables are being used to know if the variable can help increase the accuracy of a sentiment analysis system.

As finding for a variable that may help to increase the accuracy of a sentiment analyzer. The researcher upon reading literatures in emotion detection will apply the use of detected emotion in a sentence and use it to define the sentiment of a sentence. A system will be developed as a sentiment analyzer with the inclusion of emotion detection as a basis for sentiment analysis computation. The analysis will start once movie review were fed into the machine. Sentiment analysis will then begin. The parameters of this system were trained to maximize prediction accuracy given the target labels in the training set. The researcher would like to find out how will sentiment analysis enhanced by emotion detection will perform against the same sentiment analyzer machine without applying the emotion detection. once the machine is trained?

There will only be four core emotion the machine can detect in this system. The movie review as input to the machine is in Filipino(Tagalog) language. The system develop will only have four output as its emotions [1] and these are:

---

1. **Masaya** (Happiness/Joy)- the sentiment result should be positive. This emotion includes Excitement; as we are surprised or delighted with and unanticipated good thing, Gratitude; thankfulness is what emerges when one recognize that someone else's effort created a benefit for us, Pride; when a person accomplishes a goal or contribute in an important way, Serene; when there is a feeling of contentment when we find ourselves into circumstances, Interest; when we encounter something new and feel safe to explore it, and Hope; when we envision brighter future and often helps us in hard times.

2. **Malungkot** (Sadness)- the sentiment result should be negative. This emotion includes Sadness; we are being affected by the negative factors in life, Dismay; consternation and distress, typically that caused by something unexpected, Despair; feeling of complete loss or absence of hope.

3. **Pagkagalit** (Anger)- the sentiment result should be negative. This emotion includes Disgust; a strong disapproval aroused by something unpleasant or offensive, Irritation; when we feel bothered by something unexpectedly, Disappointment; when it doesn't meet our expectations.

4. **Pagkatakot** (Fear)- the sentiment result should be negative. An unpleasant emotion caused when you feel you are in danger, did something wrong, feel something wrong. The feeling of you might lose something or likely to cause pain, or a threat, Anxiety; It's a feeling of fear or apprehension about what's to come.

## **2.Related Works**

### **2.1. Movie Review**

It has been a trait of Filipino writers to impart double entendre [3] in their reviews, to a study has been made to know why many of these Filipino prolific writers impart such meanings into their pieces, one of which it imparts a better impact to the reader, said one author, she imparted a negative and positive impact to make the readers see the sides of both sides of a said movie. [3] One other reason is to make the readers dive deep into the story, to know why it was written, as an experience in itself or a lesson to be learned in the review itself.

### **2.2. Data Analysis**

Textual information can be perceived as facts or opinions, if it is a fact there are sources that backs it up and hold to be true, while when it is an opinion it is an account of what happened or based on perspective from a person. Textual information today has been categorized as to be factual data. One word in a Google Search can return many sources whether it is fact or opinion. Many authors of the said content and many critics or comments which form as opinions from the users of the internet. [4]

Movie reviews like short comments is factual or opinionated. Users tend to share mostly what they think about the said film. here we focus on two different things. People reviews can use: Subjectivity, where the sentence is there to be subjective towards the film or objective towards other things or Polarity, wherein the reviewer expresses positive or negative opinion.

### **2.3 Emotion Detection and Sentiment Analysis**

Emotion Detection and Sentiment Analysis are fields that recognition can be done through text. Two types of analysis were available to detect emotion; one is Sentiment analysis and the other is Emotion analysis. In sentiment analysis we can determine if it is positive, negative or neutral convey of emotion from a given input of text from the user, while in emotion analysis we can determine the types of emotion such as happiness, sadness, anger, disgust, fear, and surprise from given input of text from the user. [5]. A study state that people are used to express their sentiment in social media towards daily events in different life areas whether sports, political and so on. Sentiment analysis here was defined as the process of detecting sentiment or opinion of user's statement towards daily activities. [6]

Being able to grasp what emotion is being portrayed or felt through context can be very helpful. Artificial Intelligence contributes to many advancements. To implement AI, one needs to understand humans and their emotions, which in the past was considered as one of AI's limitations until accelerated access to vast data through social media and great improvement to technology. While emotion can be expressed through written words, detecting emotions through text with a computer is called Sentiment Analysis and Emotional Detection. Sentiment Analysis focuses on polarity, it classifies whether the text is positive or negative based on the sentiment information connected to the text. On the other hand, Emotion Detection is based on a wider spectrum of moods. [7]

Different tools can be used in this area like text classification. Text classification is a process of extracting textual data from reviews and classifying them into positive and negative polarity. Classification is performed by applying machine learning (ML) methods or Heuristic-based methods. The ML methods involve Support Vector

Machine (SVM), Naive Bayes (NB), and Maximum Entropy. K-Nearest neighbourhood, ID3, C5, centroid classifier, winnow classifier, and the N-gram model are the most popular machine learning methods. [6] There are several issues to be overcome in the sentiment classification. Intensive research is required on the following issues [8]: 1. Ambiguity in the text data where multiple sentiments are related to two or more issues, 2. Identification of the most effective polarity in a given document that contains both positive and negative sentiment is a major issue, and 3. Achieving successful results is tedious due to lot of noisy statements in the dataset. [8]

Learning-based strategies are being utilized to detail the issue in an unexpected way. Initially the issue was to decide emotion from information messages however now the issue is to characterize the information writings into various emotions. Dissimilar to catchphrase-based identification strategies, learning-based techniques attempt to identify emotion dependent on a recently prepared classifier, which apply different hypotheses of machine adapting, for example, bolster vector machines and contingent arbitrary fields, to figure out which emotion classification should the information content has a place. [9]

Based on some studies, analysis of sentiment and emotion is very much like data mining. They gather up reviews on some items, feedback of customers and even monitor their taste on music. What they do is they compute some of its variables and in order to understand it, apply it many times on the same subject to create a learning experience, it recognizes keywords, sentence patterns and in that way, learning and knowing, producing a good analysis on the said item. [10] For example, one study applied it on movie reviews and helped understand the emotion it evoked, not only to the readers but also the emotion of the author, it used numerous reviews on the said story gathered from people and from there data is mined and analyzed thoroughly to produce its analysis on the said review piece. [11]

Utilizing emotion keywords is a direct method to distinguish related feelings, the implications of watchwords could be various and obscure, as most words could change their implications as indicated by various utilizations and settings. Also, even the base arrangement of feeling names (without every one of their equivalent words) could have diverse feelings in some extraordinary cases, for example, unexpected or on the other hand pessimistic sentences. [5]

In sentiment analysis, there exists specific research challenges. Text Informality, Language Acronyms, Languages Mixture, Emotion icons and Relevance are samples. [6] Early works in sentiment analysis are depending on lexical resources. Preitiuc-Pietro et al. (2012) uses SentiWordNet lexicon was used and applied by counting positive and negative terms found in a review and the highest scored sentiments became the sentiment polarity for a specific review [12]. In this study construction of domain-oriented sentiment lexicon as clustering of sentiment words and extends the information-bottleneck clustering algorithm by integration more restriction for building an appropriate knowledge context of every sentiment word are challenges that were faced by the research and including also a challenge in using different tools like Opinion-Finder, WordNet-Affect, MPQA and SenticNet.

### 3. System Design And Methodology

THE SYSTEM THAT WAS DEVELOPED FOR THIS RESEARCH USED TF – IDF (TERM FREQUENCY – INVERSE DOCUMENT FREQUENCY) ALGORITHM TO TRANSFORM TEXT INTO MEANINGFUL REPRESENTATION OF NUMBERS. THIS ALGO. USES FREQUENCY (TERM FREQUENCY) INTO ACCOUNT AND PENALIZES WORDS FOR BEING OVERUSED (FOR EXAMPLE THE WORD ‘THE’ WOULD BE DOWNVOTED SINCE IT IS TOO POPULAR. THE INTUITION BEING THAT IT DOES NOT HAVE DISCRIMINATORY POWER IF THE WORD IS USED IN EVERY SENTENCE. WHEREAS WORDS MIGHT BE ABLE TO DISTINGUISH THE CONTENTS OF THE SENTENCE BETTER. THE TECHNIQUE IS WIDELY USED TO EXTRACT FEATURES ACROSS VARIOUS NLP APPLICATIONS. [13]

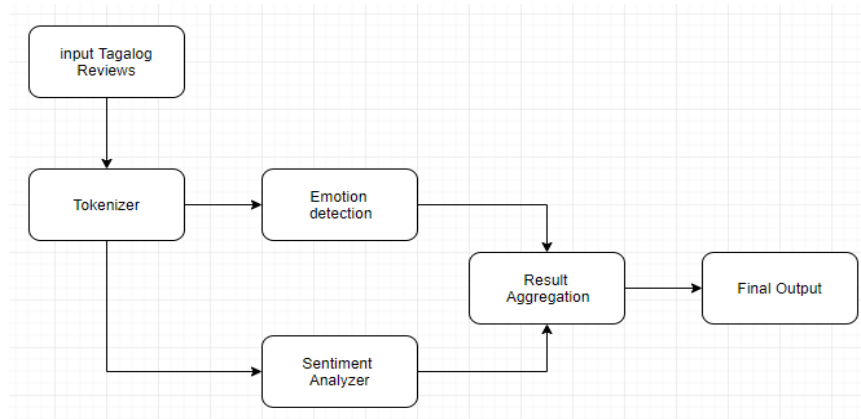


Figure 1: System Architecture

### 3.1. Text Representation

The system transformed the text into numeric or vector representation which is a must if your system will be using a machine learning algorithm. This numeric representation should depict significant characteristics of the text. There are many techniques to be used. For this system TF-IDF was used.[14]

### 3.2. Term Frequency

It is said that the higher values of word mean greater importance in a text. In some cases, it is true in some cases but might have problems if you encounter documents across corpus with different sizes. Naturally, the bigger size documents means occurrences of words than smaller documents. To minimize problem of counting word occurrences we need to normalize the occurrence of the word with the size of the corpus, this process is called term-frequency.[15]

### 3.3. Sentiment Analysis

Since there is no publicly available Filipino movie review corpus, a dataset was created. The sentiment dataset is saved in an csv format and is divided into two columns, the first column is for the Reviews and the second column is for the Sentiment label part. The dataset includes 25000 Reviews and its associated sentiment 1 means it is a positive sentiment and 0 is a negative sentiment.

	A	B		A	B
1	Reviews	Sentiment	12722	Ang mga tao, mangyaring hu	0
2	Bromwell High ay isang	1	12723	Sa totoo lang bago pinanood	0
3	Ang Homelessness (o F	1	12724	Amateur, walang badyet peli	0
4	Makinang sobrang kum	1	12725	Updated na bersyon ng isang	0
5	Ito ay madaling ang pin	1	12726	Ito ay isang tipikal na "ganap	0
6	Ito ay hindi ang karaniv	1	12727	Pagpatay ng mga numero ay i	0
7	Ito ay hindi comedic na	1	12728	Well, ang pagsusulat ay masy	0
8	Oo ito ay sining... upan	1	12729	Noong una ang pelikula ay til	0
9	Sa itong "critically ipin	1	12730	Kaya, ako 'y nagtataka habang	0
10	ANG mga TAGAPAKINIC	1	12731	Mula sa direktor Barbet Schrc	0
			12732	Alfred Hitchcock imbento ng	0

Figure 2: Data sample

Words such as ‘ako’, ‘sa’, ‘akin’, etc. that are irrelevant in figuring out what the sentiment is removed. These are commonly known as stop words in Natural Language Processing. [16] The sentiment dataset was inputted to the Tf-Idf algorithm and the algorithm will count the number of unique words and their weights and the algorithm will return a Tf-idf weighted document-term matrix. Each row in the matrix corresponds to each review and the columns corresponds to the terms of each word and the number of terms depends on the size of the vocabulary. A document-term matrix or term-document matrix is a mathematical matrix that describes the frequency of terms that occur in a collection of documents. There are various schemes for determining the value that each entry in the matrix should take. One such scheme is tf-idf. They are useful in the field of natural language processing. [14]

### 3.4. Model for Sentiment Analysis

A feed forward neural network using Keras and implemented 1 hidden layer with 100 neurons with a number of inputs equal to the total number of terms (size of vocabulary). LeakyReLU were used as an activation in the hidden layer which has been the most widely used activation function for deep learning with state-of-the-art results to date. [17] The network also has dropout layer and values of 0.2, 0.3 and 0.4 were manually tested and it was found that 0.3 gives the highest testing accuracy with 85.12% at epoch 2. [18] The output of 100 neurons is connected to the output layer consisting of only 1 neuron and is using a sigmoid activation function which are used for predicting probability-based output and has been successfully applied in binary classification problems. The test accuracy suffers overfitting for epochs beyond 4 epochs. ADAGRAD was used as an optimizer and leave the parameters at their default values. [19]

### 3.5. Emotion Recognition

Same as in sentiment analysis data, there is also no publicly available Emotion in Filipino, and data set dataset for emotion recognition was also created. The Emotion dataset is saved in an csv format and is divided into two columns, the first column is for the Labeled Emotion and the second column is for the Content. The dataset includes 3245 contents and its associated emotion.

There are only four categories of emotions to be detected. These are: ‘pagkatakot’ which means fear, ‘masaya’ means happy, ‘pagkagalit’ means anger and ‘malungkot’ which means sad. Since the dataset is small compared to the sentiment dataset, the training data used for emotion recognition is 80% of the dataset and the remaining 20% were used as testing data. We used the Tokenizer utility of Keras which allows us to vectorize the training and testing data (every character is not treated as a token) into a vector where the coefficient for each token is based on Tf-idf. Maximum number of words (size of vocabulary) is set to 2970 which is the size of the vocabulary measure by Tf-idf which is small compare to the size of vocabulary of sentiment dataset which is around 70000. Since the labels are strings instead of integers, we used a label encoder to convert label strings to numbered index and then convert the labels to a one-hot representation of both the training and testing data. The training data for the emotion recognition were 2596 content and its associated labeled emotion. The remaining 649 were used as a testing data. [20]

	A	B	C	D
1	Emotion	Content		
2	pagkatakot	ako ay naduwag		
3	malungkot	bigyan mo ako ng isa pang pagl		
4	pagkatakot	may aso!		
5	pagkagalit	AYAN! Unang una pinagtanggol		
6	masaya	10 out of 10 stars.		
7	masaya	100% irerekomenda ko 'to sa m		
8	masaya	11 over 10.		
9	masaya	Aabangan ko ung sequel.		
10	pagkatakot	aabot kaya sila		

Figure 3: Data Sample

### 3.6. Model for Emotion Recognition

After preparing the training data for the neural network, a feed forward neural network using Keras was designed and implemented hidden layer with 100 neurons with a number of inputs equal to the total number of terms (size of vocabulary which is around 2970). LeakyReLU were used as an activation in the hidden layer which has been the most widely used activation function for deep learning with state-of-the-art results to date. [17] The output of 100 neurons is connected to the output layer consisting of 4 neuron (that represents the 4 emotions and outputs the probabilities for each emotion) and is using a softmax activation function which is used in multi-class models where it returns the probabilities of each class, with the target class having the highest probability. The network also has dropout layer and values of 0.1, 0.2 and 0.3 were manually tested and it was found that 0.2 gives the highest testing accuracy with 76.57% at epoch 3. The test accuracy suffers overfitting for epochs beyond 4 epochs. [17] The researcher used “Adam” as an optimizer and leave the parameters at their default values which are based on a paper. [21]

### 3.6. Final Computation

After solving for initial sentiment and Emotion , weights were set on how to solve the final output, and then ensemble it.

Weight of Sentiment = 0.6; Weight of Emotion = 0.4

(Weight of Sentiment \* Sentiment Degree) + (Weight of Emotion \* Emotion Degree) = Final Sentiment

To understand why the generalization ability of an ensemble is usually much stronger than that of a single learner, Dietterich [22] gave three reasons by viewing the nature of machine learning as searching a hypothesis space for the most accurate hypothesis. The first reason is that the training data might not provide sufficient information for choosing a single best learner. For example, there may be many learners perform equally well on the training data set. Thus, combining these learners may be a better choice. The second reason is that the search processes of the learning algorithms might be imperfect. For example, even if there exists a unique best hypothesis, it might be difficult to achieve since running the algorithms result in sub-optimal hypotheses. Thus, ensembles can compensate for such imperfect search processes. The third reason is that the hypothesis space being searched might not contain the true target function, while ensembles can give some good approximation. For example, it is well-known that the classification boundaries of decision trees are linear segments parallel to coordinate axes.

#### 4. Discussion of the Results

CONFUSION MATRIX WAS USED TO DESCRIBE THE PERFORMANCE OF A CLASSIFICATION MODEL ON A SET OF TEST DATA FOR WHICH THE TRUE VALUES ARE KNOWN. ALL THE MEASURES EXCEPT AUC CAN BE CALCULATED BY USING LEFT MOST FOUR PARAMETERS. THESE ARE AS FOLLOWS: TRUE POSITIVES (TP) THIS STATES THAT THE SYSTEMS POLARITY OUTPUT IS POSITIVE AND THE ACTUAL POLARITY OUTPUT IS POSITIVE, TRUE NEGATIVES (TN) STATES THAT SYSTEM’S OUTPUT POLARITY IS NEGATIVE AND THE ACTUAL VALUE IS REALLY NEGATIVE , FALSE POSITIVES (FP) IMPLIES THAT THE SYSTEM’S OUTPUT POLARITY IS POSITIVE BUT THE ACTUAL POLARITY IS NEGATIVE , AND FALSE NEGATIVES (FN), IMPLIES THAT THE SYSTEM’S OUTPUT IS NEGATIVE BUT THE ACTUAL POLARITY IS POSITIVE. FOR A CLASSIFIER PERFORMANCE THE PRECISION, RECALL, AND F-MEASURE IS BEING MEASURED, AND THE ACCURACY FORMULA THAT WAS USED IS  $TP+TN/TP+FP+FN+TN$ .

The final sentiment analysis and emotion detection overall results yield the result of 46 TP, 34 TN, 8 FP and 12 FN. The results show 79.3% Recall and 85.19% Precision. The sentiment analysis and emotion detection system resulted in a high value of False Negative resulting to a low F-Measure value of 82.14%. To get the emotion detection accuracy.  $(Actual\ result - prediction\ result) / Actual\ result$ . 80% of the data set used in training the model and 20% of the remaining data set is used in testing the accuracy of the model. To get the sentiment accuracy.  $(Actual\ result - prediction\ result) / Actual\ result$ . 50% of the data set used in training the model and 50% of the remaining data set is used in testing the accuracy of the model. To get the final sentiment the percentage of the initial sentiment is multiplied by the weight with the value of 0.8 and it is added to the percentage of the emotional percentage which is added to the product of percentage of emotion to the weight of 0.2. Final Sentiment = (initial sentiment \*0.6) + (emotion percentage \* 0.4)

A system with high recall but low precision returns many results, but most of its predicted labels are incorrect when compared to the training labels. A system with high precision but low recall is just the opposite, returning very few results, but most of its predicted labels are correct when compared to the training labels. An ideal system with high precision and high recall will return many results, with all results labeled correctly.

#### 4.1. Computation of Final Sentiment

The formula used for computing the final sentiment is the weighted arithmetic mean in which it is a type of average in which each observation in the data set is multiplied by a predetermined weight before the average is calculated.

The weight assigned to Sentiment Analyzer is 0.6 and a weight of 0.4 is assigned to Emotion Recognizer. The larger dataset affects the accuracy of the model, the Sentiment Analyzer has a higher weight compared to Emotion Recognizer.

#### 4.2. Correlation of Data

Sentiment Weight	1	2	3	4	5	6	7	8	9	Mean 49.4 9%
Emotion Weight	9	8	7	6	5	4	3	2	1	
Correlation Strength	96.73%	86.67%	72.28%	57.35%	44.16%	33.38%	24.73%	17.82%	12.25%	
Accuracy	57.18%	62.46%	68.50%	76.16%	85.35%	89.13%	90.40%	90.86%	91.05%	
True Positive	11977	12131	12145	12106	11975	11843	11715	11593	11485	
True Negative	2317	3484	4981	6934	9362	10439	10885	11121	11277	
False Positive	10183	9016	7519	5566	3138	2061	1615	1379	1223	
False Negative	523	369	355	394	525	657	785	907	1015	

Correlation of Data is performed using Pearson R Test or Pearson correlation coefficient that measures the strength between variables and relationships. To measure the correlation, we ran the Emotion Detection component on the whole sentiment dataset and save all the emotion degree in an array, we also run the ensemble to the whole sentiment and save all sentiment degree in an array and ran a correlation between the two arrays to measure the correlation between the Emotion Detection and the Ensembled output. [23] Setting the weights to 0.6 to Sentiment Analyzer and 0.4 for Emotion Detection saw a high increase from 85.180% accuracy of initial sentiment to 89.128% with a correlation coefficient of 0.333784 between the final sentiment and emotion which indicates a moderately positive correlation. Weights of 0.3 to 0.1 assigned to Emotion Detection and 0.7 to 0.9 assigned to Sentiment Analyzer saw a small increase to accuracy. The mean of correlation that is gathered by collecting correlation coefficients that results from different weights assigned to Emotion and Sentiment has a value of 0.494882667 which indicates a moderately positive correlation, we can say that Emotion Detection can enhanced the accuracy of Sentiment Analyzer.

## 5. Conclusion and Future Works

This researcher was able to show that with a good data set emotion detection can help to improve the accuracy of the Sentiment Analysis. The initial sentiment prediction output is 85.1840%, after the computation and having an emotion detection the final sentiment increases by 4 to 5% The result shows that emotion plays a vital part in the accuracy of sentiment analysis because the accuracy of the mean final sentiment is 88.9245% because of the emotional value, the accuracy rises up and it determines how accurate the reading of sentiment analysis in movie reviews or Filipino language in general. According to the Rating System of Named-Entity Recognizer (NER)for Filipino Novel Excerpts using Maximum Entropy Approach. the performance of the system must be at 70% or above to say that it has a good performance. Thus, the interpretation for the system is "Satisfactory".[24]

Based on the research output it is recommended consider emoticons as variable for sentiment analysis since it was often used in expressing emotions. Combination of keywords, ending punctuation marks, and emoticons could lead into better performance in Sentiment Analysis. For future researchers, when you use the sentiment dataset in terms of emotion analyzer, use the appropriate labels of emotion instead of using the binary 1 and 0, and with that the correlation of the two will surely go higher. It is also recommended to use more tools and more analyzation to improve the accuracy of the negative sentiment, also to process the date sets to the very core and to polish it. A research to prove if sentiment analysis play a vital role in the accuracy of emotion detection can also be done. The use different activation function and more variant in order to attain a higher accuracy of sentiment analysis or a higher accuracy on learning capabilities of the program is also recommended. Furthermore, in movie review, they tend to combine two different languages. Filipino people use both Filipino and English in their tweets which mean that the language they are using is "Taglish". This is a situation which the program can't handle because its scope is on Filipino Language only.