



# Application of Naïve Bayes Algorithm in Sentiment Analysis of Filipino, English and Taglish Facebook Comments

Ronel J. Bilog

**Abstract:** *The World Wide Web has boosted its content for the past years, it has a vast amount of multimedia resources that continuously grow specifically in documentary data. One of the major contributors of documentary contents can be evidently found on the social media called Facebook. People or netizens on Facebook are actively sharing their opinion about a certain topic or posts that can be related to them or not. With the huge amount of accessible documentary data that are seen on the so-called social media, there are research trends that can be made by the researchers in the field of opinion mining. A netizen's comment on a particular post can either be a negative or a positive one. This study will discuss the opinion or comment of a netizen whether it is positive or negative or how she/he feels about a specific topic posted on Facebook; this is can be measured by the use of Sentiment Analysis. The combination of the Natural Language Processing and the analytics in textual form is also known as Sentiment Analysis that is use to the extraction of data in a useful manner. This study will be based on the product reviews of Filipinos in Filipino, English and Taglish (mixed Filipino and English) languages. To categorize a comment effectively, the Naïve Bayes Algorithm was implemented to the developed web system.*

**Keywords:** *English, Filipino, Naïve Bayes Algorithm, Sentiment Analysis, Social Media, Taglish.*

## I. INTRODUCTION

The Philippines, for third year in a row, has appeared as the social media capital in the whole world, as millions of Filipinos devote hours on social media sites, predominantly Facebook, the uppermost social media that has been used in the world [1]. Based on the 2018 Global Digital statement, Filipinos consumed a typical of three (3) hours and fifty-seven (57) minutes on social media on previous year. They are known to be the first technology adopters and remain internet savvy, particularly when it comes in communicating online. In addition to the wide-ranging usage of the social media, Filipinos based on the report said that they spent an average of nine (9) hours and twenty-nine (29) minutes on the internet each day [2].

Statements are very influential tool inside social media. When consumer is examining for a product, they depend on

the words and understandings of others to support them in making their decision. Businesses or some organization can get reviews and convert them into an advertising creative. The influence of reviews on social media is very strong. Businesses can possibly take every chance to gather comments or reviews and can use the information in their social strategy.

Customers or consumers are now greatly affected in creating buying decisions by the reviews or comments in the online websites similar to Yelp and Trip Advisor, this is based on the study that was conducted by the research firm Dimensional Research. Based on the positive reviews of products that were appeared online, there is ninety percent (90%) of customers that their decision on purchasing products were affected by online comments after reading it. On the other hand, negative reviews and comments also affected the consumers on buying products about eighty-six percent (86%) after reading it based on the study that involved one thousand forty-six (1,046) respondents.

Most negative reviews and comments were commonly found on websites that are dealing with product reviews, while the social media such as Facebook was gained forty-four percent (44%) of positive comments was came from it based on the survey [3].

Sentiment analysis or opinion mining is contextual mining of text or data which classifies and extracts particular info in source material, and serving a business to recognize the social sentimentality of their brands, products or even services while monitoring online discussions. Moreover, to gain the feedback of customers with regards to a specific topic by the advertisers or any other organization, the sentiment analysis is being used.

This research intends to analyses an opinion, review or comment to determine whether it is positive, negative or neutral using the sentiment analysis by applying Naïve Bayes algorithm in Filipino and English languages.

## II. AIM OF THE PAPER

The objective of this research is to classify an opinion or comment coming from the social networking site Facebook based on sentiment analysis whether it is positive or negative by the used of sentence-level classification.

Specifically, it aims to:

1) apply the Naive Bayes algorithm in web sentiment analysis; and

Revised Manuscript Received on January 03, 2020.

\* Correspondence Author

**Ronel J. Bilog\***, Junior Lecturer, Department of Computer Education, City College of Calamba, Calamba City, Philippines. Email: ronnel.bilog@gmail.com

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an [open access](https://creativecommons.org/licenses/by-nc-nd/4.0/) article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

2) classify a comment expressed in Filipino, English and Taglish languages through Facebook according to their states (positive or negative).

### III. RELATED STUDIES

#### A. Naïve Bayes Algorithm

The Naïve Bayes algorithm verifies to be the utmost effective among three (3) algorithms (SentiWordNet, Naïve Bayes and Logistic Regression) for classification of text of opinion mining. Based on the study conducted by Majumdar et al. [4], they used three (3) different algorithms to study the reviews of customers from the Amazon website. Their research can expand by using additional algorithms to classify reviews coming from different websites by mining the product reviews of customers [4].

The performance of classification approaches can be created out by using the following metrics: (1) recall is the count of true positive and describes as the proportion of correct instances that classified as the given class over the number of actual total in that class; (2) precision describes as the ratio of the correctly classified over number of the all experimental classifications; while (3) F-Measure is the combined portion for precision and recall. Based on the researchers, among the three (3) algorithm Naïve Bayes algorithm has fairly better outcomes over the SentiWordNet and Logistic Regression techniques [4].

In the paper of Troussas et al. [6], they presented the sentiment analysis for linguistic learning with the use of Naïve Bayes Classifier. Additionally, they presented significant features that reach an important gain over a unigram baseline. In addition to their study, they explored a different technique of data representation also report significant developments over the unigram models. The researchers' data is random sample of updated Facebook statuses and were not gathered by using precise queries. The size of their hand-labeled data permits them to perform the cross-validation experimentations and checked for the variance in the performance of classifier across folds. Lastly, they presented their experimental results, which showed that the accurateness in evaluating the sentiment state of Facebook users, using Naïve Bayes Classifier is really high [6].

#### B. Sentiment Analysis

Sentiment Analysis denotes to the application of the Natural Language Processing, computational linguistics and the text analytics to classify and extract subjective data in the source materials such as discussions with regards to certain products or services. In the study of Patacsil et al. [5], they used sentiment analysis to the gathered data from comments section of particular blogs. Blogs are commonly one of the many platforms that shows personal opinions in an exact topic, and comments section such as blogs commonly contains the opinions and feedbacks of the readers of a website blog. Certainly, comment section holds numerous types of expressions that can be either negative or positive in nature, and these can possibly be indicative of the customer approval [5].

Based on the study of Lapitan et al. [8], the analysis of emotions automatically carried in content in social media,

such as tweets, has numerous beneficial applications. One of the greatest disaster-prone nations in the world is the Philippines, such methods could possibly enable primary responders to make appropriate decisions in spite of the danger of data deluge. However, recognizing emotions stated in the Philippine-generated tweets, which are typically written in English, Filipino or mix of both, is non-trivial job. In order to simplify the change of natural language processing procedures that will systematize such kind of analysis, the researchers created corpus of tweets which predominant emotions were manually interpreted by the used of crowdsourcing. Defining measures confirming that only first-rate annotations were maintained, they produced a gold standard of corpus of one thousand one hundred forty-six (1,146) English and Filipino emotion-labelled tweets. They validated the value of the manually formed resource by representing an automatic emotion-prediction process based on the use of openly accessible word-emotion association lexicon was incapable to duplicate the labels allotted via crowdsourcing.

#### C. The Filipino Language

Filipino is the certified language of the Philippine country. Based on the Article XIV Section 6 of the 1987 Philippine Constitution, —the National language of the Philippines is Filipino. As it changes, it shall be more developed and improved on the foundation of present Philippine and other languages. Filipino is primarily created on the Tagalog dialect. The constitution also delivers provision for the usage of English as given the by law. The provision to practice English was drawn from the impact of American profession of the Philippines. The implementation of English and Filipino caused the Filipinos to be bilingualism. Since Filipino language appeared to had other traces of language effects of Mexicans, Chinese, Sanskrit, and Arabs, the presence of English added more difficulty to the Filipino language [7][11][12].

However, before the end of Marcos administration, a new language was progressed from Tagalog and English. The Tagalog dialect and English language were mixed to produce semantically-acceptable sentences. This was the start of Taglish (TAGalog + EngLISH). “Life is short, gawin mo ang mga mabubuting bagay na magpapasaya sa’yo.” is a sample of Taglish sentence that can be interpreted in English as “Life is short, do the good things that will make you happy.” The handiness of mixing English and Tagalog in the Philippine country is very prominent. It has been detected by linguists that nowadays, Taglish is frequently encountered in the Internet commonly in social networking sites [7][12][13].

### IV. METHODOLOGY

#### A. Sentiment Analysis Methodology

This paper focuses on mining comments or reviews from the social networking site Facebook.com, which permits user to post a status update and to write a review or comment in a post without restrictions.



It uses algorithm such as Naïve Bayes algorithm to categorize the review as negative, positive or neutral review. The Fig.1. illustrates the data flow of the proposed system. The different processing mechanisms of the system are as follows:

- Positive and Negative Comments  
This module contains the negative and positive word lists which will be saved in the two (2) separate text file and later it will be used for the sentiment analysis.
- Training Data  
The sentences were further processed to remove stop words (Common words that have a little value in the “a”, “ay” “the,”ang”) and stemming (words that carry similar meanings, but in different grammatical forms such as “runner”, and “running” was combine into one word “run”) was applied also in the sentence preprocessing.
- Naive Bayes Classifier  
Naïve Bayes classifier algorithm is created on applying Bayes’ theorem [9, 10]. Document analysis and text classification are around of the applications of this set of rules. The equation for Bayes’ theorem and its algorithm is given on the below Formula 1 and Figure 1, separately.

$$\text{posterior probability} = \frac{\text{class prior probability} \times \text{likelihood}}{\text{predictor prior probability}} \quad (1)$$

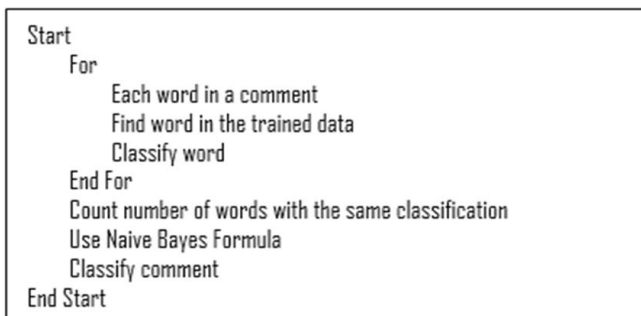


Figure 1. Naive Bayes’ Algorithm [7].

The algorithm revealed in Figure 1 will primary categorize all the words in the Facebook comment based on the different affect states. This will be done using the word bank. When the classification is completed, the Naïve Bayes algorithm will use the formula to categorize the comment. The formula will multiply the number of times the classification occurred in the Facebook comment (class prior probability) and the probability of observing the classification in the Facebook comment (likelihood). This will be divided to the predictor prior probability.

- Facebook comments  
This function contains the Facebook comments from the different users which will be used as an input data for the sentiment analysis. Comments were selected based on the content of a social media post; in this study, the researcher used e-commerce page as an input data.
- Display the Classification Result  
This module displays the results of the processed reviews lists containing positive and negative review separately.

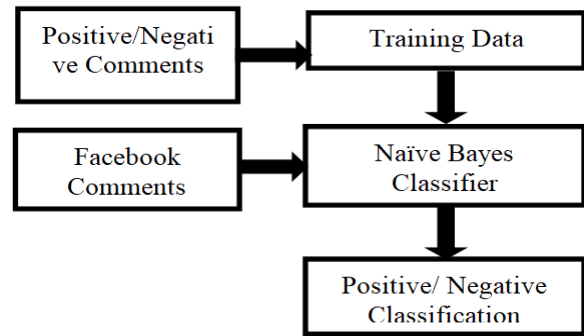


Figure 2. Methodology of Sentiment Analysis [6].

Only Facebook comments posted by Filipinos were gathered. Since Filipinos are bilingual, both English and Filipino Facebook comments were measured. Data gathering were directed within a span of one month. There were 300,000 Facebook comments collected with regards to the product reviews.

### B. Measures and Analysis

The effectiveness of the web sentiment analysis was evaluated based on accuracy, precision, recall and F-score. Accuracy is the total correctness of the classification (See Formula 2.). An accuracy of one-hundred percent (100%) means that the predicted instances are precisely the same as the actual instances. On the other hand, precision is the accuracy of the classification on each class or category. This is described in Formula 3. Table 1 shows the confusion matrix for the derivation of accuracy and precision.

Table I: Confusion matrix

	Predicted positives	Predicted negatives
Actual positive instances	Count of True Positive instances (TP)	Count of False Negative instances (FN)
Actual negative instances	Count of False Positive instances (FP)	Count of True Negative instances (TN)

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \quad (2)$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (3)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (4)$$

$$\text{F-score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (5)$$

The portion of true positive instances that were predicted over the positive actual instances is called recall while the average of precision and recall is known as F-score.



**Table II: Effectiveness rating scale**

Percent (%) Range	Verbal Interpretation
81.00%-100.00%	Highly Effective/ Highly Accurate/ Highly Precise
61.00%-80.00%	Effective/ Accurate/ Precise
41.00%-60.00%	Moderately Effective/ Moderately Accurate/ Moderately Precise
21.00%-40.00%	Slightly Effective/ Slightly Accurate/ Slightly Precise
00.00%-20.00%	Not Effective/ Not Accurate/ Not Precise

The 1000 sample Facebook comments were randomly picked. To verbally determine the numerical findings of the result per metric, the above effectiveness rating scale was used.

**V. RESULTS AND DISCUSSION**

In this research, the researcher used the Naïve Bayes sentiment classifier to classify sentences polarity and also used in predicting whether a Facebook comment is positive or negative.

**Table III: Sample positive and negative comments**

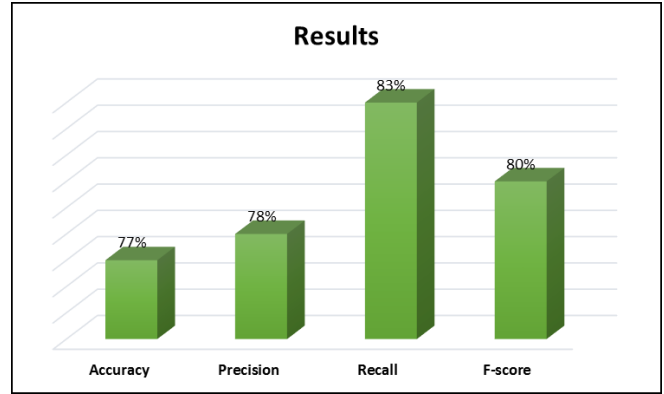
Sample Positive Comments	Sample Negative Comments
hahaha ganda nung red divaa?	Fake daw yan!
Wow	Nkakalungkot lng may nkita na akong a10s na fake at benibenta na sa market place ng fb.
Hi, Have a Great Day! Beautiful Posts! Thanks	Badtrip?? wala pa nag halfyear may A10S na. Sayang ung A10?????
eto oh 😊	Rip A5s
not bad as well mas cheaper ni	Bt... this aint no A10. 😞

To determine Facebook comments whether it is positive or negative, it was then manually labeled. To classify the unlabeled sentences or comments based on its polarity, the Naïve Bayes classifier was then trained. The example of Facebook comments on each class was shown in table 3.

**Table IV: Data results**

	Predicted positives	Predicted negatives
Actual positive instances	475	100
Actual negative instances	135	290

The web sentiment analysis classified the 290 negative Facebook comments correctly over the 390 instances of negative comments, and out of 610 total positive comments, there were 475 comments that were correctly classified as shown in table 4 data results.



**Figure 3. Result per metric.**

The figure shows that the developed web sentiment analysis has 77% of accuracy with verbal interpretation of “accurate”. Out of 1000 test data, there are cases of 765 in total that were classified accurately. It was revealed that the precision and recall of negative Facebook comments have better results over the positive ones. Therefore, the classification of sentiment polarity on comments of the web sentiment analysis was proficiently achieved.

**VI. CONCLUSION**

This study attempted to categorize comments expressed by Filipinos in Facebook whether it is positive or negative. A web sentiment analysis was used to attain this goal. It was exposed that the web system was effective in distinguishing positive and negative comments. On the other hand, the web sentiment analysis was found and classify as precise, effective, and accurate based on the overall result. The precision and accuracy of the web sentiment analysis can still be increase and enhance since it also has boundaries. Case sensitivity when distinguishing words can be avoided. To further improve the web sentiment analysis, the scanning analysis should be upgraded.

As a recommendation to the web system, the trained data should have a collection of comprehensive netizens’ terminologies from chats like “wer na u, ma peyk”. Last of all, the sample size of the Facebook comments should be increase for further analysis.

**REFERENCES**

- Pablo, M.C., 2018. Internet Inaccessibility Plagues “Social Media Capital of the World” [Online]. Available from: <https://asiafoundation.org/2018/10/24/internet-inaccessibility-plagues-social-media-capital-of-the-world/>
- Mateo, J., 2018. “Philippines still world’s social media capital – study.” Available from: <https://www.philstar.com/headlines/2018/02/03/1784052/philippines-still-worlds-social-media-capital-study>
- Pinkham, R., 2019. “90% of Consumers Say Online Reviews Impact Buying Decisions ... And Other Hot Topics.” Available from: <https://blogs.constantcontact.com/do-customers-trust-online-reviews/>
- Kumar K L, S., Desai, J., Majumdar, J. 2016. “Opinion Mining and Sentiment Analysis on Online Customer Review.” Nitte Meenakshi Institute of Technology, Bangalore, India. DOI: 10.1109/ICCIC.2016.7919584
- Patacsil, F., Malicdem, A., Fernandez, P., 2015. “Estimating Filipino ISPs Customer Satisfaction Using Sentiment Analysis.” Computer Science and Information Technology 3(1): 8-13. DOI: 10.13189/csit.2015.030102



6. Troussas, C., Virvou, M., Junshean Espinosa, K., Llaguno, K., Caro, J., 2013. "Sentiment analysis of Facebook statuses using Naive Bayes Classifier for language learning." 4. 1-6. 10.1109/IISA.2013.6623713.
7. Pippin Jr., M., Odasco, R.J., De Jesus Jr., R., Tolentino, M.A., Bringula, R., 2015. "Classifications of Emotion Expressed by Filipinos through Tweets." International MultiConference of Engineers and Computer Scientists Vol I. ISBN: 978-988-19253-2-9.
8. Lapitan, F.R., Navarro, R., Albacea, E., 2016. "Crowdsourcing-based Annotation of Emotions in Filipino and English Tweets." Proceedings of the 6th Workshop on South and Southeast Asian Natural Language Processing, pages 74–82.
9. Shawkat Ali, A.B.M., Wasimi, S.A., 2007. "Data mining: Methods and Techniques." Thomson Learning: Australia, pp.96-103
10. Han, J., Kamber, M., 2006. "Data mining: Concepts and techniques, 2nd edition." Morgan Kaufmann Publishers: San Francisco CA, USA.
11. The 1987 Philippine Constitution. Retrieved from: <http://www.gov.ph/constitutions/the-1987-constitution-of-the-republic-of-the-philippines/the-1987-constitution-of-the-republic-of-the-philippines-article-xiv/>.
12. Bugayong, L.K., 2011. Taglish and the social role of code switching in the Philippines, Philippine Journal of Linguistics, vol. 42, pp. 1-19.
13. Thompson, R. M., 2003. Filipino English and Taglish. John Benjamins Publishing Co.: Amsterdam.

### AUTHOR'S PROFILE



**Ronel J. Bilog**, finished his Bachelor's Degree in Information Technology at Laguna State Polytechnic University – San Pablo City Campus (LSPU-SPCC) and his Master's Degree in Information Technology at Rizal Technological University (RTU) in Mandaluyong City. Presently, he is taking his Doctorate Degree in Information Technology at AMA University (AMAU) in Quezon City, Philippines. He is a former Information Technology Instructor at LSPU-SPCC and Laguna University in

Sta. Cruz, Laguna. He is an active member of Philippine Society of Information Technology Educators (PSITE). He is a current Junior Lecturer at City College of Calamba (CCC), Calamba City. His research interest focuses on developmental research in technology, computer science, data management and data mining.