*Available online a*t www.scientiaresearchlibrary.com

Scientia Research Library **ISSN 2348-0424**
USA CODEN: JETRB4

Journal of Engineering And Technology Research, 2018, 6 (3):1-6

(*http://www.scientiaresearchlibrary.com/arhcive.php*)

ESTIMATING CUSTOMER SATISFACTION USING NAÏVE BAYES MACHINE LEARNING MODEL

Joseph B. Campit
Information Technology Department,
Pangasinan State University – Bayambang Campus, Philippines
Corresponding Author *:* *josephbcampit@gmail.com*

ABSTRACT

The study investigated the classification performance of Naive Bayes (NB) Machine Learning Model in estimating customer satisfaction. Likewise, it also determined the effect of applying different model configurations such as n-gram, stop words, and stemming on the classification performance of the model. Sentiment analysis was employed to analyze useful information from the unstructured responses of the respondents. The dataset consists of Tagalog and English words that were manually annotated and were randomly selected and assigned to the training and testing dataset. The general framework of the study consists of data preprocessing, modelling, and model comparison. Finding revealed that the highest classification performance of NB is attained using NB trigram with stop words removal.

Keywords : Machine Learning, Naïve Bayes, Sentiment Analysis, Text Analytics, classification performance

INTRODUCTION

The amount of unstructured data available is tremendously increasing at an exponential rate. Roughly 80% of the information in an organization is stored in unstructured formats [1]. With the increasing availability of unstructured documents, the task of automatic retrieval and classification of useful knowledge from large volumes of human-generated text is very important for several situations.

Sentiment analysis is a growing area of interest in Natural Language Processing (NLP) using machine learning models. It has become an emerging research method in various fields, including education [2], economics [3], social sciences [4], bioinformatics [5], business [6], engineering [7], and marketing [8]. It involves classifying opinions in text into categories like "positive" or "negative" or "neutral". Sentiment analysis can be defined as a process that automates the mining of attitudes, opinions, views, and emotions from text, speech, tweets, and database sources through Natural Language Processing. It's also referred to as subjectivity analysis, opinion mining, and appraisal extraction [9].

In this study, the Naïve Bayes (NB) machine learning model was experimented on by applying different model configurations such as n-gram, stop words, and stemming using a customer satisfaction data. NB is a simple probabilistic classifier based on applying Bayes' Theorem and is used for solving sentiment classification problems [10].

General Framework of the Study

The general framework of the study consists of data preprocessing, modelling and model comparison. Figure 1 shows the general framework of the study that was used to determine the best model in estimating customer satisfaction.

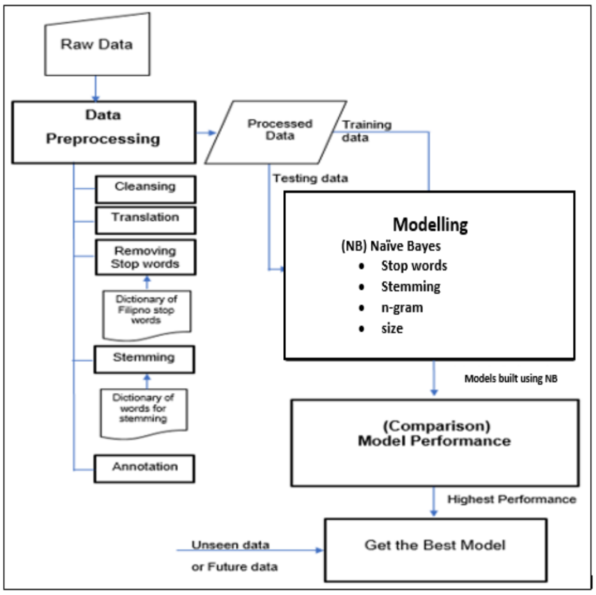


Figure 1: General Framework of the Study

Data Preprocessing

In this study, the written responses of the respondents were manually encoded in MS Excel. During the data pre-processing, data cleansing was done by removing irrelevant words, and characters in the sentence. English words were also translated to Filipino since not all responses were written in Filipino. The Google Translate tool was utilized to automatically convert English words into Filipino.

Stop words were also removed in the pre-processing of data such as *"ay", "sa", "at", "ang", "mga",* etc. These are the common words in Filipino that have little value in the process of identifying category of the documents. A dictionary of Filipino stop words was created and was used to remove all stop words from the document.

Stemming was also applied in the data pre-processing. This was done to convert different word form into similar canonical form or to return the word to its stem or root. This means that words with the same canonical form is counted as one such as *"magaling", "pinakamagaling", "napakagaling"* since they all have the same canonical form *"galing"*. In the stemming process, the stemmer will remove all the affixes (prefix, infix, and suffix) and the reduplicated parts and retrieve the root word only. Affixes in Filipino includes *"um", "ma"* and *"in"*. Words with these affixes include *"b(um)ilis, (ma)ayos, s(in)abi.* After stemming these words, *"bilis", "ayos"* and *"sabi"* will be retrieved respectively. In the words *"aangat"*, and "tataas", the morphene *"a-"* and "*ta*-" was reduplicated. The purpose of stemming is to reduce the words in the document and achieve smaller processing time. After stemming, the morphene "*a-*" and *"ta-"* will be removed and *"angat"* and "taas" will be retrieved. A dictionary of Filipino words or affixes for stemming was created to stem the words in the document.

The identification of sentiment polarity was manually annotated. Seven groups with three members each were tasked to manually annotate the polarity of the sentences. The groups of raters were oriented and trained.

The Fleiss' Kappa inter-rater reliability test was employed to test the rater's consistency of the annotation process. Based on the results, all the kappa values in each group are within acceptable inter-rater reliability (k>0.75) for applied test. The mean inter-rater reliability for the sentence polarity raters was found to be k = 0.79. According to the definition of the Fleiss' Kappa statistic, the accuracy of the inter-rater reliability is considered to be "Substantial agreement" [11],[12].

Modeling

In this stage, the text in the document is represented in a form that can be used to train the algorithm. The main task in this training phase is to assign a label to the newly encountered documents from the pre-defined categories. *The Naive Bayes* classifier was trained and tested using the training and testing data. The application of different model configurations such as the use of n-grams, removing stop words, and stemming were employed in classifiers training and testing and then, evaluated the performances to build a proposed machine learning model. The 10-fold cross-validation was used to validate the classification performance of the classifier.

The proposed models built using the NB were compared to determine the model with the highest performance. Based on the evaluation results, the model with the highest performance is the best or recommended machine learning model.

OBJECTIVES OF THE STUDY

This study investigated the classification performance of Naive Bayes Machine Learning Model in estimating customer satisfaction.

Specifically, it determined the classification performance of NB machine learning model parameterized by varying values of: n-gram, stop words, and stemming. Lastly, it determined the machine learning model that gave the best result in estimating customer satisfaction.

MATERIALS AND METHOD

Research Design

This study used quantitative method of research. Sentiment analysis was employed to classify positive and negative sentiments from the unstructured responses of the respondents in a customer satisfaction survey.

Respondents

The respondents of the study were the 8,000 randomly selected personnel and students of Pangasinan State University during the 2nd semester of the school year 2017 - 2018.

Dataset

The dataset of the study consists of 56,000 sentences and was divided into two: the training dataset and the testing dataset. The training dataset was used to train the machine learning classifier while the testing dataset was used to evaluate the effectiveness of the built machine learning models. The

75% of the dataset was used for training and 25% was used for testing. Random sampling was employed in selecting the training and testing dataset of the study using the random function (Bernoulli) with an initial seed of 122,714.

Statistical Treatment of data

To ensure the reliability of the results, RapidMiner 8.2 (free) Basic Edition was used. It provides all the steps of the text mining process such as data preprocessing, visualization of results and validation.

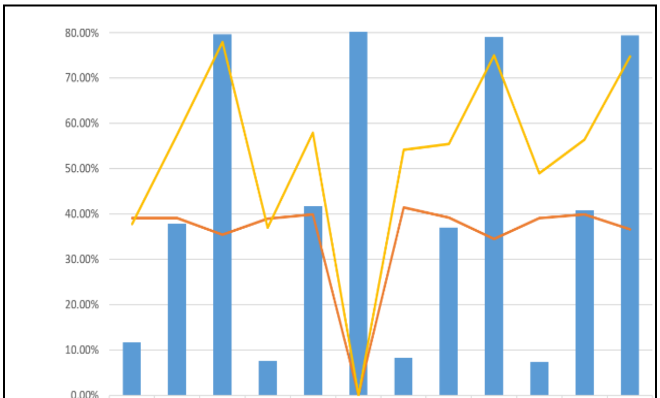
To determine the classification performance of Naive Bayes machine learning model parameterized by varying values of size, n-gram, stop words, and stemming, the generated values *in the confusion matrix* were used in the computation of the classification accuracy, precision, recall and F-measure.

To determine the best model in estimating customer satisfaction, a comparison of their performance was done.

RESULTS AND DISCUSSION

CLASSIFICATION PERFORMANCE OF NB

Figure 2 presents the classification performance of NB applying different combinations of n-gram, stop words, and stemming.



	n-gram			n-gram + stopword			n-gram + stemming			n-gram + stopword + stemming		
	1-gram	2-gram	3-gram	1-gram	2-gram	3-gram	1-gram	2-gram	3-gram	1-gram	2-gram	3-gram
Accuracy	11.72%	37.91%	79.61%	7.58%	41.71%	80.20%	8.34%	37.07%	79.02%	7.38%	40.78%	79.44%
F-measure (pos)	39.16%	39.14%	35.42%	38.96%	39.91%	0.00%	41.45%	39.24%	34.50%	39.14%	39.91%	36.65%
F-measure (neg)	37.86%	57.57%	77.96%	36.98%	57.95%	0.00%	54.17%	55.44%	74.97%	49.03%	56.40%	74.81%

Figure 2. Classification Performance of NB

Figure 2 reveals that in terms of n-gram, NB trigram obtained the highest classification accuracy of 79.44%. In terms of n-gram with the application of removing stop words, NB trigram obtained the highest classification accuracy of 80.20%. Furthermore, the result of n-gram with the application of stemming words shows that NB trigram obtained the highest classification accuracy of 70.02%. Moreover, in terms of the combination of n-gram, removing stop words, and stemming, NB trigram obtained the highest classification accuracy of 79.44%. This indicates that NB trigram attained the highest classification accuracy in all combinations.

It is important to mention that the classification performance of NB increases from unigram to trigram within each of the combinations such as n-gram, removing of stop words, stemming, and the combination removing stop words and stemming. These ranges from 11.72% to 79.61%, 7.58% to 80.20%, 8.34% to 79.02%, and 7.38% to 97.44% respectively. This implies that bigram and trigram words are better n-gram representations to be used to improve the performance of NB than unigram.

The results show that the best classification performance of NB is obtained using NB trigram with stop words removal achieving a classification accuracy of 80.20%. On the other hand, the lowest classification result is obtained using NB unigram with stop words removal and stemming obtaining a classification accuracy of only 7.38%. The figure also shows that the classification performance of NB trigram in all combinations of model configuration features is high ranging from 79.02% to 80.20%. However, the classification performance of NB unigram and bigram is low in all combinations ranging from 7.38% to 41.71%. This observation also agrees with the study of Patacsil [13] which emphasized that n-gram works well on the classification of sentiments and showed that trigram is effective

CONCLUSION

NB had a high classification performance in trigram applying stop words removal, stemming, and combination of stop words removal and stemming. However, NB had a low classification performance in unigram and bigram applying all combinations of model features. The highest classification performance of NB in estimating customer satisfaction using unstructured data in Filipino words is attained using NB trigram with stop words removal.

REFERENCE

[1] Raghavan, (**2004**). S. Amer-Yahia and L. Gravano eds., “Structure in Text: Extraction and Exploitation.” In. Proceeding of the 7th International Workshop on the Web and Databases (WebDB), ACMSIGMOD/PODS 2004, ACM Press, Vol 67, 200

[2] Lin, Hsieh, & Chuang, (**2009**). Discovering genres of online discussion threads via text mining. Computers and Education, 52, 481-495

[3] D Zubin,J., Bruce,K., and Suresh, N. (**2014**). Detecting latent ideology in expert text: Evidence from academic papers in economics. In Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), pages 1804–1809. Association for Computational Linguistics.

[4] Wilkerson, J., Smith, D., and Stramp, N. (**2015**). Tracing the flow of policy ideas in legislatures: A text reuse approach. American Journal of Political Science, pages n/a–n/a.

[5] Consoli, D. (**2009**). Analyzing customer opinions with text mining algorithms. AIP Conference Proceedings, 1148, 857-860.

[6] Devitt A, Ahmad K. **2007**. Sentiment polarity identification in financial news: a cohesion-based approach. In Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics,Vol. 7, Prague, Czech Republic,1–8

[7] Kostoff, R. N., Karpouzian, G., & Malpohl, G. (**2005**). Text mining the global abruptwing-stall literature. Journal of Aircraft, 42, 661-664.

[8] Gans et al., (**2017**). Exit, tweets and loyalty. NBER Working Paper 23046, National Bureau of Economic Research, Camb ridge MA. Available at: htpp://www.nber.org/papres/w23046.

[9] Vishal A. Kharde, S.S. Sonawane, (**2016**). International Journal of Computer Applications (0975 – 8887) Volume 139 – No.11.

[10] Medhat et. Al, (**2014**). Sentiment Analysis Algorithms and Applications: a Survey. Ain Shams

University. Ain Shams Engineering Journal 5, pp. 1093-1113

[11] Ling, Y. How to Conduct Inter-rater Reliability Tests?. Overseas Chinese Association for Institutional Research An AIR Affiliate That Supports IR Professionals Since **1996**

[12] Polit, D. F. and Beck, C. T., Nursing Research: Generating and Assessing Evidence for Nursing Practice, 8th edition, Lippincott Williams & Wilkins. **2008**

[13] Patacsil, (2015). Blog comments Sentence Level Sentiment Analysis for Estimating Filipino ISP Customer Satisfaction. International Conference Data Mining, Civil and Mechanical Engineering (ICDMCME '2015) February 1-2, **2015** Bali (Indonesia)