Analyzing Sentiment of Myanmar Movie Comments Using Naïve Bayes Classifier

Win Win Thant, Ei Ei Mon
University of Computer Studies, Hinthada
winwinwthant@gmail.com, eemucsy@gmail.com

***Abstract*— Sentiment analysis is extremely useful in social media monitoring as it allows us to gain an overview of the wider public opinion behind certain topics. It is also the use of Natural Language Processing (NLP) to analyze social comments and determine a deeper context. Due to little research has been done in the domain of Facebook sentiment classification about Myanmar movie comments, this system aims to determine the overall reaction of the movie using Naïve Bayesian machine learning technique. The objective of this system is to extract movie features from the comments, classify the comments into positive and negative and predict the overall polarity of the comments. In the data preparation phase, the system collects movie comments written by Myanmar language in Facebook as a dataset and these comments are first cleaned to reduce noise. Text pre-processing is also done on Myanmar Word Segmentation Version 1.0 of UCSY and movie features are extracted. Classification is made by using Naïve Bayes approach that gives rise to a probabilistic classifier. For evaluation, 10-fold cross validation which is a standard machine learning assessment technique has been performed on a balanced data set of the movie comments. And the performance of the proposed approach was also evaluated using F-measure.**

Keywords— Sentiment Analysis; Naïve Bayes; Facebook; Myanmar Movie Comments;

I. INTRODUCTION

The growing expansion of contents, placed on the Web, provides a huge collection of textual resources. People share their experiences, opinions or simply talk just about whatever concerns them online. A large amount of available data attracts system developers, studying automatic mining and analysis [1]. In this system, Facebook considered as a rich resource to collect people's opinions in the movie domain and attracted researchers to develop an automatic Sentiment Analysis (SA) model for comments of the post. These days there are so many "critics" online and audiences' feedback analysis is one of the essential components for improving the movie. Because of the massive volume of comments, audiences can't read all comments.

The overall polarity of the comments is important factor to make a decision for some audiences whether or not it is recommended to watch a certain movie or not and for the entire cast and crew of a movie to enhance or upgrade the quality of their movie. The producers can collect the user's opinion whether favorable or not about their movie and then they can enhance and upgrade the quality of their movie as quality improvement. By analyzing and categorizing the people's opinion according to their preferences and interests, the system can predict which movie should be recommended and which one should not be recommended. The result of sentiment analysis techniques can be used in marketing research. By this technique, the attitude of viewers about some movies can be analyzed.

The key contributions of this work are:

- Extending of the previously created dataset of movie domain
- Appending of positive and negative opinion words to previously proposed opinion lexicon
- Extraction of movie features by preprocessing steps
- Classification of Facebook comments into positive and negative opinionated sentences by using Naïve Classifier
- Estimating the polarity of the comments

To our knowledge, no prior work had been done exclusively on the analysis of polarity on Facebook Myanmar movie comments until only recently.

For the sake of convenience the remainder of this paper is organized as follows: Section 2 presents the related works of the system that are machine learning approaches for sentiment classification. Section 3 introduces the method of the system which includes the data sources, opinion words and the task of classification using Naïve Bayes approach. The fourth section is about the performance evaluation done on the system. The fifth section discusses about the challenges of the system. Last section concludes our study and discusses some future directions for research.

II. RELATED WORK

Much work has recently been undertaken in sentiment analysis over the last few years.

A. Sentiment Analysis using Twitter

This part of the paper is used to explain the related study of sentiment analysis on twitter.

A study by M. Vadivukarassi et al. [2] collected the keywords from Twitter using Twitter API and preprocessed the extracted raw data using Natural Language Toolkit (NLTK) techniques. The sentiments of the online tweets are evaluated based on feature selection of score words. Chi Square test is used to select the best features and Naïve Bayes classifier is used for training and testing the features. They conducted the experiments based on different features such as 10,100,1000,10000 respectively. They gained an average accuracy of 0.821 for all the number of features in the tweets by using Naïve Bayes classifier approach. Their proposed system would be easy for user to obtain summarized report about the opinion from Twitter.

By considering a framework, the authors [3] implemented Sentimentor, a web based tool which uses

naive Bayes Classifier to classify live Twitter data based on positivity, negativity and objectivity. Sentimentor has an interface which enables the user to analyze the word distributions. It presents classification results in an easy to understand pictorial format and includes the text type details, the analysis of the twitter message and search. Twitter API was used for the data extraction process. The OpenNLP library was used for POS tagging and the extraction of unigrams and bigrams. 216 manually tagged tweets are used in their training set and POS tags, unigrams or bigrams are considered as features in their experiments. They compared results in different n-grams of POS tags.

A study conducted by Parul and Teng-Shen [4] performed text mining on 42,235 tweets collected from five national political parties in India to predict election results. They used Twitter Archiver tool to get tweets in Hindi language. They made use of both supervised and unsupervised approaches and utilized three distinct strategies Dictionary Based, Naive Bayes and SVM algorithm to build the classifier and classified the test data as positive, negative and neutral. They concluded that SVM performed the best amongst various others.

B. Sentiment Analysis using Facebook

In this section, we now look at some related study of sentiment analysis on Facebook.

A study by G. Vashisht and S. Thakur [5] created manually emoticon sentiment lexicon. They demonstrated how emoticons typically convey sentiments and how they can exploit emoticons. They applied finite state machines to find out the polarity of the sentence or paragraph. The experiment is based on 1,250 Facebook status and 2,050 Facebook comments, which all contain emoticons. The most commonly and frequently used emoticons were identified in their system. They classified them on the basis of the sentiment they strengthen which eventually decides the polarity of the sentence. Their system can be used together with a lexicon-based sentiment analysis method to validate the result.

The authors [6] proposed to mine opinion of Thai people about the current government revolution. The opinion was extracted from the Facebook statuses updates which were written in Thai language. They implemented the system with two feature extraction methods. First, they used the traditional pre-process of text mining to extract the features. Second, they constructed a sentiment lexicon by collecting positive and negative words to extract features. Comparative experiments were performed among Naïve Bayes, Support Vector Machines (SVM), K-Nearest-Neighbor (KNN) and decision trees. Their experimental results showed that KNN gives the highest accuracy. It provided 63.58% of accuracy for mining opinions of Thai politics.

By considering a framework, the researchers [7] discussed a methodology which allows utilization and interpretation of Facebook data to determine public opinions. They classified their reviews into relevant and non-relevant using Naïve-Bayes classifier. Analysis was done on reviews about the android. Ratings of various mobiles based on user reviews from Facebook are explained. In their experiments, they showed count of rating by product name, processing speed, battery backup and OS Version such as Kitkat, Lollipop, Marshmarllow. They stated that Kitkat is more preferable then other versions of OS for android phones.

C. Sentiment Analysis on Reviews

Sentiment analysis of reviews such as music reviews, movie reviews, product reviews, hotel reviews etc. is considered to be very challenging and the related study of sentiment analysis on reviews are explained in this section.

This paper [8] presented an empirical study of efficacy of classifying product review by semantic meaning. The authors proposed hybrid algorithm combining Decision Trees and Naïve Bayes algorithm to classify the polarity of comments given on e-commerce websites. They used a web crawler to fetch comment on a particular web page. Word Net dictionary is used for the spelling correction to make the most sensible comment for knowing the polarity of words. After classifying the positive and negative words using Naïve Bayes algorithm, the overall polarity is calculated using decision tree.

The researchers [9] proposed a comparative study of the effectiveness of ensemble technique for sentiment classification. Naive Bayes (NB) and Genetic Algorithm (GA) are combined to be a new hybrid classification method with the aim of efficiently integrating different feature sets and classification algorithms to synthesize a more accurate classification procedure. The advantages of their approaches are demonstrated by means of movie reviews. The methodology is based on five main parts: pre-processing phase, document indexing phase, feature reduction phase, classification phase and combining phase to aggregate the best classification results. The results are evaluated using the cross validation method. They concluded that the hybrid classifier shows the significant improvement over the single classifiers.

The problem of sentiment polarity categorization which is a fundamental problem of sentiment analysis is tackled in this paper [10]. Online product reviews collected form Amazon.com are use as dataset. Experiments for both sentence-level categorization and review-level categorization have been performed. They used scikit-learn as software and Naïve Bayesian, Random Forest, and Support Vector Machine as classification models. They observed that the Random Forest model performed the best for sentence-level categorization. Both the SVM model and the Naïve Bayesian model were identical in terms of their performances for review level categorization and both were superior to the Random Forest model.

III. PROPOSED METHODOLOGY

People's sentiments, ideas, feelings are a very important factor to make a decision. While seeing any movie, user's first to read the opinions and comments of that particular movie and those reviews have a great impact on the user's mind. Social Network Sites (SNSs) play important roles in people's lives for sharing information. Facebook becomes one of the important platforms for interaction. Facebook allows people to have their own accounts to comment, express feelings and convey emotions via texts as well as emoticons [11].

In the proposed system, the movie comments are extended from the previous dataset and these comments are categorized into positive and negative using opinion words. The dataset is cleaned in the preprocessing step and the classification is done based on the preprocessed movie features to define the polarity of comments. The system architecture is shown in Fig. 1.

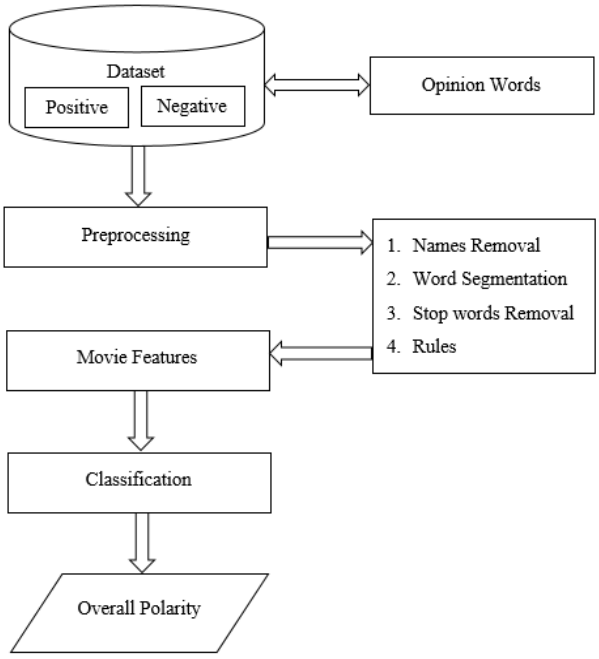

Figure 1. Proposed System

Our proposed methodology includes the following steps:

A. *Extending Dataset*

The dataset used in SA are important. The main sources of data are from the user-created movie comments written by Myanmar language in Facebook. These comments are important to the business holders as they can take business decisions according to the analysis results of users' opinions about their movies. Most people want to know which movie is good and they trust the recommendations from friends. This system presents an extended version of dataset by adding new comments to an existing dataset [12] collected from Facebook, from November to March 2016. We show the links of newly added comments to the existing dataset and the example of randomly selected movie data is shown in Table 1. For example, 1,350 comments are collected from the Facebook movie link of bridgeofclouds (တိမ်တိုက်တံတား).

TABLE I. TOTAL NUMBER OF COMMENTS FOR SOME MOVIES

Facebook Movie Web Pages Link	No. of Comments
https://www.facebook.com/bridgeofclouds/	1350
https://www.facebook.com/shwekyarmovie/	750
https://www.facebook.com/weretigerfilm/	450
https://www.facebook.com/MSMMMOVIE/	1200
https://www.facebook.com/DaddyPharKim.Movie/	900

B. *Appending Opinion Words*

Our approach depends on the existence of previously proposed opinion words [12] that provide information about 44 positive words and 35 negative words for the movie domain. However, there are few features for the movie domain. So, we define an extensive set of 30 opinion words for positive and 25 words for negative. The extensive opinion words are derived manually through the analysis of subjective words that frequently occur in the comment of the newly added dataset. Some positive and negative opinion words are shown in Table 2.

TABLE II. OPINION WORDS

Positive features	Negative features
ကောင်း (good)	ရပ် (stop)
အားပေး (encourage)	ပျင်း (lazy)
အောင်မြင် (succeed)	စိတ်ပျက် (disappointed)
ကြိုက် (like)	ပေါ (silly)
လေးစား (respect)	ဝေဖန် (criticize)

C. *Classifying Positive and Negative Comments*

The comments were classified into two types positive and negative comments based on extensive set of opinion words. If the positive opinion word count is greater than the negative count, it is classified as positive; otherwise, it is classified as negative. 2000 positive comments and 2000 negative comments are used in this system and example comments are shown in Table 3.

TABLE III. POSITIVE AND NEGATIVE MOVIE COMMENTS

Type	Comments
Positive	ကိုမြင့်မြတ်နဲ့ကိုခင်လှိုင်ဆိုကြိုက်ပြီးသားအမြဲအားပေးနေတယ် သရုပ်ဆောင်တာအရမ်းပီပြင်တယ် အကြိုက်ဆုံးဇာတ်ကားပဲ အကယ်ဒမီရပါစေလို့ညီမဆုတောင်းပေးပါတယ် ကြည့်ပြီးပြီ အရမ်းသဘောကျတယ်
Negative	ငပေါမှတကယ့်ငပေါကား ဟာသကားတဲ့မရယ်ရတဲ့အပြင်ကြည့်ရင်းနဲ့ကိုပျင်းလာတယ် အခြောက်ကားတွေပဲရိုက်တက်ကြတော့တယ် မြန်မာရုပ်ရှင် စိတ်ပျက်ဖို့ကောင်း သရုပ်ဆောင်တွေ မကြိုက်ဘူး

D. *Preprocessing*

The dataset has to be cleaned because the writing on social networking websites is normally not correct. Therefore, some words in a dataset are changed to the correct form to improve efficiency and effectiveness. The rules in previous work are also used for the data set. Besides, the steps included the removal of any names of actor and actress. And then segmentation process is done on Myanmar Word Segmentation Version 1.0 of UCSY [13]. Some rules are added to the segmentation result in order to implement in this system. Two segmented words are combined to be one word. For example, မ_ကြိုက် to မကြိုက် (do not like), ပေါ_ကား to ပေါကား (silly movie). Some of the spelling errors are manually corrected. Finally, a set of words these are not important in our system are removed from the comments. Their appearance in a comment does not give any valuable data and some useless words are shown in Table 4.

TABLE IV. USELESS WORDS

Types	Words
Particles	ပြီးပြီ, တယ်, ပေါ့, လေ , ပါ, ခဲ့ , နော်
Pronouns	သူတို့ , ငါ , သူ , ခင်ဗျား
Postpositional markers	သည် , က , မှာ, အတွက် ,လို
Conjunctions	လျှင် , ကို , နှင့်, သော
Numerical Classifiers	နှစ်ယောက်, တစ်ကား
Adverbs	လုံးဝ , အကုန် , အရမ်း ,တော်တော် ,အသေ

The steps of preprocessed movie features extracted from our dataset are shown in Fig. 2.

Sample comment

မြင့်မြတ်သရုပ်ဆောင်ကောင်းတယ်ဇာတ်ကားကို ကြိုက်တယ်

After name removal

သရုပ်ဆောင်ကောင်းတယ်ဇာတ်ကားကို ကြိုက်တယ်

After word segmentation

သရုပ်ဆောင်_ကောင်း_တယ်_ ဇာတ်ကား_ကို_ကြိုက်_တယ်

After useless word removal

သရုပ်ဆောင်_ကောင်း_ဇာတ်ကား_ကြိုက်

Figure 2. Steps of Preprocessing

E. Dataset of Movie Features

A feature or aspect is an attribute or component of an entity. After preprocessing, the movie features dataset becomes something like this and is shown in Table 5.

TABLE V. POSITIVE AND NEGATIVE PREPROCESSED DATA

Type	Preprocessed Movie Features
Positive	ကြိုက်_အားပေး သရုပ်ဆောင်_ပီပြင် ကြိုက်_ဇာတ်ကား အကယ်ဒမီရ_ဆုတောင်း_ ကြည့်_သဘောကျ
Negative	ငပေါ_ငပေါကား ဟာသကား_မရယ်_ကြည့်_ပျင်း အခြောက်ကား_ရိုက် မြန်မာ_ရုပ်ရှင်_စိတ်ပျက် သရုပ်ဆောင်_မကြိုက်

F. Polarity Classification by Naïve Bayes

This is a binary classification task with two classes of sentiment polarity: positive and negative. It is made by using Naive-Bayes approach, i.e., a machine learning algorithm that gives rise to a probabilistic classifier, which works on the basis of the Bayes Theorem, with the strong assumption that features are mutually independent. Naive Bayes classifier assumes that the presence of a particular feature in a class is unrelated to the presence of any other feature. This model is easy to build and particularly useful for very large datasets. Along with simplicity, Naive Bayes is known to outperform even highly sophisticated classification methods. The basic idea of Naive Bayes technique is to find the probabilities of classes assigned to texts by using the joint probabilities of words and classes and the classifier is shown in Fig. 3.

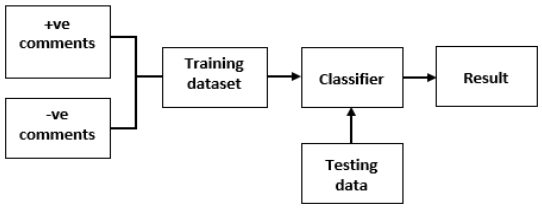

Figure 3. Naïve Bayes Classifier

Let us note c_i is one of the classes (positive or negative) and f_j is the one of the features in a comment. Under a probabilistic framework, the comment is then assigned to the class with the highest probability, i.e.

$$c^* = \arg\max_i P(c_i \mid f_j) \quad (1)$$

By using the Bayesian rule, the posterior probability can be decomposed into:

$$P(c_i \mid f_j) = P(f_j \mid c_i) * P(c_i) / P(f_j) \quad (2)$$

Moreover, since the denominator is the same for both classes we can leave it out here. So, we have:

$$c^* = \arg\max_i P(f_j \mid c_i) * P(c_i) \quad (3)$$

Finally the two probabilities to estimate on the training data are the prior probability of a class $P(c_i)$ and the likelihood $P(f_j|c_i)$.

The prior probability is simply estimated by counting the number of comments of each class:

$$P(ci) = \frac{\text{no of comments in class ci}}{\text{no of all comments}} \quad (4)$$

Because we make the assumption that the words are independent, we can write:

$$P(f_j \mid c_i) \text{ as } P(f_1|c_i)*P(f_2*c_i)*\ldots..*P(f_j*c_i) \quad (5)$$

The likelihood is simply estimated by counting the number of features of each class, and by additive smoothing not to put a disadvantage on less populated classes:

$$P(fj|ci) = \frac{\text{no of fj occurs in ci} + 1}{\text{no of comments in class ci}} \quad (6)$$

A small example using the Naïve Bayes in given dataset below:

	f1	f2	f3	f4	f5	f6	f7	f8	class
	မင်းသား	ကြိုက်	ဇာတ်ကား	ကောင်း	ပေါကား	ရုပ်	အခြောက်	အားပေး	-
S1		yes		yes	yes			yes	pos
S2			yes			yes	yes		neg
S3			yes		yes	yes			neg
S4	yes	yes	yes	yes					pos
S5		yes	yes				yes		pos
S6	yes		yes		yes		yes		neg

S1= ပေါ်ကား ပေမယ့် ကောင်းတယ်ကြိုက်တယ်အားပေးမယ်

P(S1|pos) = P(pos)*(ပေါ်ကား |pos) *(ကောင်း |pos) *(ကြိုက်|pos) *
(အားပေး |pos)
= 3/6 * (1+1)/3 *(2+1)/3 *(3+1)/3 *(1+1)/3 = 0.2963

P(S1|neg) =P(neg)*(ပေါ်ကား |neg) *(ကောင်း |neg) *(ကြိုက်|neg) *
(အားပေး |neg)
=3/6 * (3+1)/3 *(1+1)/3 *(0+1)/3 *(0+1)/3 = 0.0494

S1 is positive comment because 0.2963 > 0.0494.

S2= အခြောက်ဇာတ်ကားများရပ်

P(S2|pos) = P(pos)* P(အခြောက် |pos)* P(ဇာတ်ကား |pos)* P(ရပ် |pos)
=3/6 * (1+1)/3 *(2+1)/3 *(0+1)/3 = 0.1111

P(S2|neg) = P(neg)* P(အခြောက် |pos)* P(ဇာတ်ကား |neg)*
P(ရပ်|neg)
=3/6 * (2+1)/3 *(2+1)/3 *(2+1)/3 =0.5

S2 is negative comment because 0.1111<0.5.

G. *Analyzing the overall polarity*

In this process, sentiments are extracted from entire movie comments. The goal is to classify whole movie comments as positive or negative. After determining the polarity of each comment, the overall polarity can be calculated. If the positive comments are greater than the negative comments for a movie, the overall polarity is positive and vice versa.

IV. EVALUATION

Performance of proposed technique is evaluated with our collected dataset (Section 3) and for this purpose cross validation (CV) and three performance parameters are used: Precision, Recall and F-Measure.

A. *Cross Validation*

Generally speaking, over-fitting happens when the training data is relative small, and CV is a good solution to avoid this. K-fold CV represents the K number of folds/ subsets. Our training set is further split into k subsets where we train on k-1 and test on the subset that is held. This is done for each k fold with a k scores given as a result. We average the model against each of the folds to finalize our model. By rotating through the subsets of training data it helps the resulting model to generalize.

For evaluation, 10-fold cross validation which is standard machine learning assessment technique has been performed on a balanced data set of the movie comments. The data set is divided into 10 subsets of equal size and the overall accuracy result for 10-fold cross validation is 83.6% and is shown in Table 6.

TABLE VI. CROSS VALIDATION RESULT

Round	1	2	3	4	5	6	7	8	9	10
Accuracy	83 %	85 %	84 %	84 %	81 %	85 %	84 %	80 %	87 %	83 %

B. *Performance Method*

The performance of the proposed approach was evaluated using F-measure. It is the harmonic mean of precision and recall. Precision and recall are two standard evaluation metrics widely used to evaluate the effectiveness of classification algorithms on a given category [14]. Table 7 presents a confusion matrix for binary classification, where tp are true positive, fp – false positive, fn – false negative, and tn – true negative counts.

TABLE VII. A CONFUSION MATRIX FOR BINARY CLASSIFICATION

Class	**As positive**	**As Negative**
Positive	tp	fn
Negative	fp	tn

The precision (P) can be calculated using tp and fp rate as in (7):

$$P = \frac{tp}{tp + fp} \tag{7}$$

tp is used for the sentences which are correctly classified whereas fp is used for sentences, which are wrongly classified.

The recall (R) can be calculated as in (8):

$$R = \frac{tp}{tp + fn} \tag{8}$$

fn is used for non-classified sentences and tp are the sentences, which are correctly classified (as explained above).

Given P and R, the F-measure (F) is defined as in (9).

$$F = 2.\frac{P.R}{P + R} \tag{9}$$

For evaluation purpose, different numbers of comments collecting from Facebook are used as a test set. There are about 300 positive comments and 300 negative comments in the test data set. The results of Precision, Recall and F-measure are shown in Fig. 4 and Table 8.

TABLE VIII. EXPERIMENT RESULTS

Class	**As positive**	**As Negative**	**Precision**	**Recall**	**F-measure**
Positive	242	58	0.776	0.807	0.791
Negative	70	230	0.799	0.767	0.783

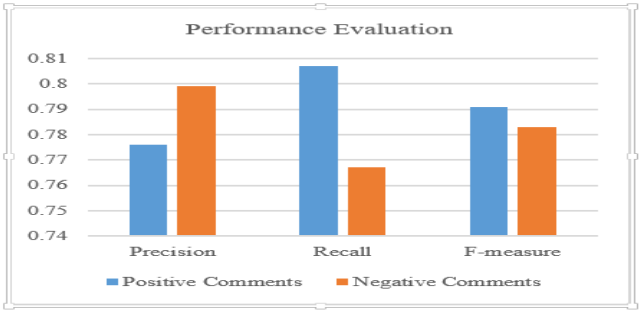

Figure 4. Performance of the system

V. DISCUSSION

Sentiment analysis is commonly used by movie creators and other organizations that wish to acquire their customers' reaction on a specific topic. Despite naïve Bayes technique is very simple and easy to implement, it still holds some issues on this system.

Data collection is a basic step for any sentiment analysis. Benchmark data set are not available free for Myanmar language in sentiment analysis field. So, we can collect a small amount of data.

The final outcome of data pre-processing will help for better sentiment analysis. But the pre-processing steps for Myanmar language is very challenging. The writer does not follow any constraints and uses a different writing style. These challenges become obstacles in analyzing the exact meaning of sentiments and detecting the suitable sentiment polarity. Sometimes, we manually corrected the data.

Although this system obtains the good result on our dataset, its main drawback is that when a new feature comes, the system should be re-trained in the classification of positive and negative comments.

VI. CONCLUSION

User comments on movies on the web are an important information source in entertainment. Sentiment analysis plays a vital role to make a decision like movie comments. We present a system that collects such comments from the web and classifies them. The classification of Myanmar comments is more challenging than in other, more richly resourced languages. Classification of comments in both positive and negative classes is accomplished based on a Naive Bayes algorithm. The system achieved an overall classification accuracy of 83.6% for 10-fold cross validation on the dataset of 4000 movie comments. The result of F-measures for positive and negative comments are 0.791 and 0.783 respectively.

In future, we will be finding out the best result of sentiment analysis by developing more data of movie comments, more opinion words, and more data cleaning rules. We will try to build a comprehensive SA system for Myanmar movie comments. One of our future works is to experiment with different classifiers on our dataset. Word sense disambiguation will be considered to get high accuracy.

REFERENCES

[1] C. Troussas et al., "Sentiment analysis of Facebook statuses using Naive Bayes classifier for language learning" , In: 2013 Fourth International Conference, Information, Intelligence, Systems and Applications (IISA), pp.1-6, IEEE (2013)

[2] M. Vadivukarassi, N. Puviarasan, and P. Aruna, "Sentimental Analysis of Tweets Using Naive Bayes Algorithm", World Applied Sciences Journal 35 (1): 54-59, 2017, ISSN 1818-4952

[3] J. Spencer, and G. Uchyigit, "Sentimentor: Sentiment Analysis of Twitter Data", In: Proceedings of European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases, pp. 56-66, 2012.

[4] P. Sharma, and T. Moh, "Prediction of Indian Election Using Sentiment Analysis on Hindi Twitter", In: 2016 IEEE International Conference on Big Data (Big Data), pp. 1966-1971, 2016.

[5] G. Vashisht and S. Thakur, " Facebook as a Corpus for Emoticons-Based Sentiment Analysis", International Journal of Emerging Technology and Advanced Engineering (IJETAE), ISSN 2250-2459, ISO 9001:2008 Certified Journal, Vol. 4, Issue 5, May, 2014.

[6] P. Songram and C. Jareanpon, "Opinion Mining of Thai Politics on Facebook Status Updates", Proceedings of the 3rd IIAE International Conference on Intelligent Systems and Image Processing 2015 (ICISIP 2015)

[7] K. Srividya et al., "Sentiment analysis of facebook data using naïve bayes classifier", International Journal of Computer Science and Information Security (IJCSIS), Vol. 15, No. 1, January 2017.

[8] G. Kaur and A. Singla, "Sentimental Analysis of Flipkart reviews using Naïve Bayes and Decision Tree algorithm", International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) ,Vol. 5, Issue 1, January, 2016.

[9] M. Govindarajan, "Sentiment Analysis of Movie Reviews using Hybrid Method of Naive Bayes and Genetic Algorithm", International Journal of Advanced Computer Research (ISSN (print): 2249-7277 ISSN (online): 2277-7970) Vol. 3, No. 4 Issue13, December, 2013.

[10] X. Fang and J. Zhan, "Sentiment analysis using product review data", In: Journal of Big Data (2015) 2: 5. doi:10.1186/s40537-015-0015-2

[11] N. A. M. Zamani et al., "Sentiment Analysis: Determining People's Emotions in Facebook", In: Proceedings of the 13th International Conference on Applied Computer and Applied Computational Science. April, 2014.

[12] W. W. Thant and K. Shirai, "Automatic Acquisition of Opinion Words from Myanmar Facebook Movie Comments", In: Proceedings of the LREC 2018 Workshop "The 13th Workshop on Asian Language Resources", isbn: 979-10-95546-24-5, pp. 75-81, Miyazaki, Japan, May 7, 2018.

[13] http://www.nlpresearch-ucsy.edu.mm/NLP_UCSY/wsandpos.html

[14] https://en.wikipedia.org/wiki/Precision_and_recall

[15] https://en.wiktionary.org/wiki/Category:Burmese_particles

[16] https://en.wikipedia.org/wiki/Burmese_numerical_classifiers