Sentiment Analysis of User Feedback in e-Learning Environment

Mohd Asri Omar[1], Mokhairi Makhtar[2], Mohd Fauzi Ibrahim[3], Azwa Abdul Aziz[2]

[1,3]*Faculty of Informatics Science, University College Bestari, Setiu, Terengganu, Malaysia*
[2]*Faculty of Informatics and Computing, Sultan Zainal Abidin University, Besut, Terengganu, Malaysia*

ABSTRACT
Sentiment analysis (SA) is prevalent now; because it can yield useful insight from high-volume subjects and unstructured data mainly from social media networking sites and micro-blog websites or known as user-generated contents. SA in user-generated contents is difficult due to the informal nature of the communication. The informal nature introduces additional variables and properties that have to be evaluated compared to formal texts, necessitating additional resources spent on annotating the data and training the classifiers. We explore two most common methods in classifying user-generated contents called lexicon-based approach by using VADER Sentiment Analyser and Machine Learning (ML) approaches by using Naïve Bayes and Decision Tree classifiers. Our primary objective is to study the accuracy of the solutions and then apply the best solution to 126 students' feedbacks 126 student feedbacks toward an e-learning environment. The purpose is to extract the sentiment against it to acquire the initial picture of student's perception on the implementation of e-learning; so the effectiveness of its implementation can be improved. The data pre-processed and then analysed using Python as programming tool. The results show VADER outperformed two selected ML classifiers, it can achieve approximately 90% in accuracy. From the results, we conclude that VADER sentiment analyser was doing well and better than ML in SA toward user-generated content. The results on the e-learning environment also suggest further analysis should be done towards this e-learning platform to complete this initial study.

Keywords : *Sentiment analysis, lexicon-based, machine learning and e-learning.*

I. INTRODUCTION

Sentiment Analysis (SA) is useful to a wide area of problems that are good to human-computer interaction practitioners and researchers, as well as those from various disciplines such as sociology, marketing, and advertising, psychology, economics, and political science [1] . It needs a lot of information coming from different sources and about various topics to be retrieved and fused [2] .

In learning analytics, SA used to measure, analyse, report and predict data about learners to optimize teaching and learning. The data that usually get examined are structured data including grades, attendance data, login frequency and site participation concerning a learning management system and so on. Data analysis is leading to a user experience modelling to determine whether a learner is satisfied with the learning experience. Besides that, it is vital to understand the patterns produced by data like student feedback to effectively improve the performance of the institution and to create plans to enhance institutions' teaching and learning experience [3].

Sentiment classification in user-generated contents is difficult due to the informal nature of the communication. The informal nature introduces additional variables and properties that have to be evaluated compared to formal texts, necessitating additional resources spent on annotating the data and training the classifiers [4] .

In this study, SA is performed to choose the best classifier in classifying the user-generated content, then it will be used to analyse the user feedback on e-Learning environment. The purpose is to extract the sentiment against it to acquire the initial picture of student's perception on the implementation of e-learning so the effectiveness of its implementation can be improved.

II. SENTIMENT ANALYSIS APPROACHES

There are two major approaches used in sentiment analysis: lexicon-based approach and ML approach. The lexicon-based approach begins with the pre-processing of the text to be analysed. The total score is initialized to zero. Then it is checked the presence of the lexicon in the dictionary of words and determined if the lexicon is present in the dictionary of words and if present whether it is positive or negative and the score is updated accordingly. The final score will thus classify the text as positive or negative [5].

ML approach relies on the famous ML algorithms to solve the SA as an established text analysis problem that makes use of syntactic and linguistic features. ML strategies work by training an algorithm with a training data set before applying it to the actual data set. An ML technique first trains the algorithm with some particular inputs with known outputs so that later it can work with new unknown data [6]. A study summarise the approaches used in SA like shown in Figure 1 [7].

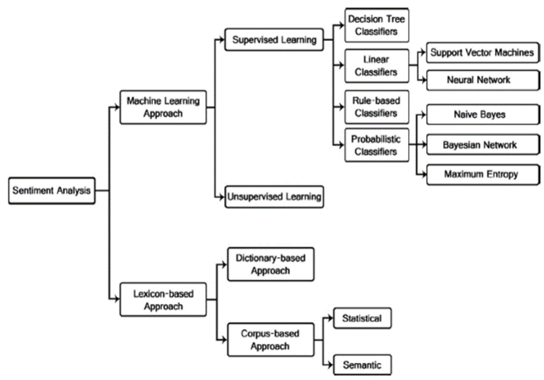

Figure 1: SA Classifiers

III. PROPOSED METHOD FOR SENTIMENT ANALYSIS

This section discusses the method proposed which involves three phases starting from data collection and integration, data transformation, and sentiment extraction. Figure 2 below summarised the research methodology process for conducting this research.

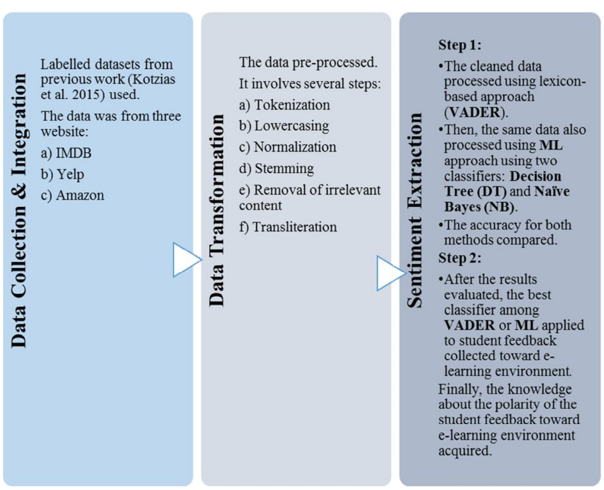

Figure 2: Research Methodology Process

A. Data Collection and Integration

The dataset was chosen because of its relevance to the subject of sentiment classification for informal short text generated by customer. The dataset from previous work done by other researchers [8].

This dataset contains sentences labelled with positive or negative sentiment only. It has already been manually annotated and this eliminates the need for distant supervision and the baseline to which the results are compared will be highly reliable. This dataset was created for the paper From Group to Individual Labels using Deep Features. The sentences come from three different websites: IMDB.com (IMDB), Amazon.com (Amazon) and Yelp.com (Yelp).

For each website, there are exist 500 positive and 500 negative sentences. We use three datasets (IMDB, Yelp and Amazon) from previous work to test the classifiers, the reason is the datasets are also user-generated content and are similar in characteristics as feedbacks collected from students.

a) Data Transformation

Data pre-processing is the process of cleaning and preparing the text for classification. Texts contain usually lots of noise and uninformative parts. This step could be a good method for normalizing messages or removing the noise before applying the SA. In most cases, these methods are used for removing or changing some parts of the messages. Pre-processing processes can be simple or very complex and are a great help, since they can be used in a non-intrusively and independent way. These pre-processing systems can be used always, for all messages, or eventually. They depend on the current trends or the dynamic evolution of the messages topic.

During this phase, we prepared the data for classifying process. In this study, we are also not comparing the accuracy results before and after pre-processing. The previous study done [9] proved that text pre-processing in sentiment analysis may be significantly improved the results. Other study also emphasize on the pre-processing phase [10,11,12,13].

Usually, pre-processing involves six (6) steps (not all steps run for each selected classifier such as stop words is not removed when using VADER Analyser). Pre-processing is conducted according to the needs of particular approach.

1) Tokenization: Comments are split into sentence, or tokens using tokenize function in NLTK. Then, the message is divided into sentences. Dots are the only punctuation marks considered as separators at this step, since others such as commas or semicolons can be part of emoticons.
2) Lowercasing: Characters are converted to lower case to ease the process of classifying the text.
3) Normalization: Abbreviated content is normalized by using a dictionary to map the content to frequently used Internet slang words. For example, "gud" and "awsm" are mapped to "good" and "awesome," respectively. Words that describe sentiment but are not in lexical-based dictionary are also converted so that matching with words in the dictionary. It is translated into synonyms or

related words. For example, "slow" converted to "low".

4) Stemming: The purpose of stemming process is to facilitate word matching. In this process, words in student comments are converted to their root word.
5) Removal of irrelevant content: Punctuation and stop words (stop words are words that are often useless in NL Processing), which are irrelevant for sentiment analysis, are removed to improve system response time and effectiveness.
6) Transliteration: To address the issue of use of mixed language in student comments, the text is transliterated using the Google Transliterate API. By using Text Blobs, the sentences can be translated between languages. If no source language is specified, TextBlob will attempt to detect the language. Language translation and detection are powered by the Google Translate API.

B. Sentiment Extraction

In this phase, each feedback will be processed using two commonly used methods in sentiment analysis; ML and Lexicon-based. For ML, two classifiers selected; Naïve Bayes and Decision Tree. For Lexicon-based approach, VADER has been selected based on the extraordinary performance review that has been made.

The main objective of this phase is to extract the informal short text generated by user to know whether it is positive, negative or neutral statement. In the first step, the performance between ML and VADER will be compared

IV. RESULTS AND DISCUSSION

The results showed VADER can achieve approximately 90% in accuracy when evaluated all selected data set (IMDB, Amazon and Yelp). It is the same as claimed by previous researchers [1] that VADER Sentiment Analyser performs a very satisfying performance for Sentiment Analysis. It can achieve an excellent accuracy for every dataset. Table I below displays the results when the datasets processed by VADER Sentiment Analysers using 0.05 as threshold value as recommended by VADER documentation.

The same datasets were used to evaluate ML techniques. Randomly, 700 data used to train the classifier and then 300 data used to test it. The full evaluation results for Naïve Bayes and Decision Tree also shown in Table I below. NB classifier can achieve 81% accuracy in Amazon dataset, 74% in IMDB dataset and 78% in Yelp dataset. The performance of DT classifier quite low compared to NB. This classifier achieves 76% in Amazon dataset, 63% in IMDB dataset and 69% in Yelp dataset.

Table I: Results comparison among ML and VADER

	Accuracy (%)			
Dataset	**Naïve Bayes**	**Decision Tree**	**VADER** (with 0.05 threshold value)	**Total reviews**
Amazon	81	75.67	90.91	1000
IMDB	74	62.67	89.66	1000
Yelp	78.33	69.33	90.63	1000

The results show that VADER outperformed both of the selected ML classifiers: NB and DT. Therefore, the Lexicon-based approach: VADER Analysers then applied to extract the sentiment from user feedback on the selected e-learning environment.

A. Sentiment Analysis on e-Learning Platform.

VADER has been used to extract the sentiment from 126 students. Before that, the comments were pre-processed, and for that purpose, not all steps involved in pre-processing in ML were used. For example in this case stop words and symbols have not been removed because they are useful in VADER as it was explained in VADER documentation. Besides that, the uppercase in the reviews also were not changed as it is valuable for VADER.

Then, the review from a student may be fragmented to two, three or more sentences because in sentence's tokenization process every sentence was separated by dots. Dots are the only punctuation marks considered as separators. So, the total of sentences becomes one hundred thirty-eight (138) reviews.

After been processed, VADER classified the reviews to sixty-three (63) positive feedbacks, thirty-five (35) neutral feedbacks, and thirty-nine (39) negative feedbacks. It means that the platform got more positive reviews than negative reviews from VADER. But the neutral sentiment was quite high. So, it must be looked closely why it happened, and from our observation, it was due to the absence of sentiment word in the sentences especially which described as the negative sentiment.

To resolve the issue, we normalize three (3) words that describe as negative, but they are not in the VADER dictionary. The slow, small and down converted to low, low and no response. After normalization process, VADER classified just nine sentences as neutral as the result shown in table II below. It did not solve all the issues in this study but managed to reduce the number of sentences that were not successfully classified to acceptable amounts. According to the SA results done using VADER Analyser in Table II and Figure 2 below, 51% of the comments are negative, 42% are positive, and 7 % are neutral.

It is clear that the management should take the lead by conducting a follow-up study by doing an aspect-level analysis to find out what is the real issue behind it to ensure that the problems are adequately addressed. The most important thing, it will benefit the lecturers and student as the users of this platform.

Table II: Classification Results after Normalization

Polarity	Positive	Neutral	Negative
Total	58	9	71

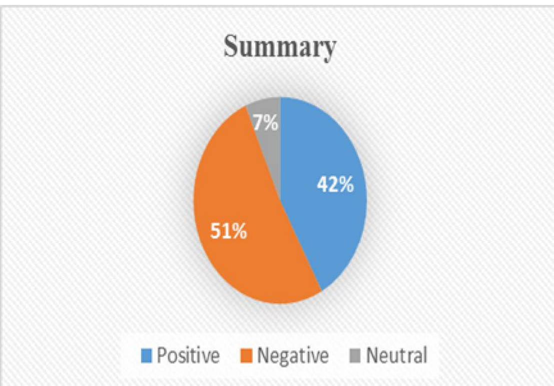

Figure 3: Classification Results after Normalization

To confirm whether the comments provided by the student represent their overall perceptions on the e-learning platform or just represent their opinions on the particular components in that platform, we ask for their overall rating in the same survey and the result shows that the students who participate in this survey were satisfied with the e-learning environment. While in the earlier analysis against it shows a lot of negative comments, this does not mean they are totally dissatisfied with it and should be emphasized that these two results are in fact not contradictory to each other because the comments given by them are only touching what they are not satisfied with. So this is the reason why we propose an aspect-based study implemented to find out what is the real issue behind the negative comments.

In the rating result, eighty-five (85) students gave positive rating (good and excellent) to this e-learning platform, only four (4) students gave their overall negative feedback (poor) and thirty-six (36) students gave neutral rating (average) on this platform. Their rating score is as in Table III.

Table III: Rating Results

Rating	Total
Excellent	11
Good	74
Average	36
Poor	4
Very Poor	0

So it can be said that this platform succeeded in helping the teaching and learning process in this institution. Nevertheless, it is still widely open for further improvements, so further analysis on the negative comments provided by the students should be made and it should not be neglected by the management.

V. CONCLUSION

Because SA is a branch of data mining and text mining. So, it must follow the five steps in the Knowledge Discovery in Database (KDD) process. It comprises steps like data collection, text pre-processing, sentiment detection and sentiment classification. There are two main methods used in sentiment analysis, namely ML and lexical-based. No technique can be considered best in all times and circumstances. It depends on the domain and issues that the researcher wants to analyse or tackle.

VADER Sentiment Analyser can perform well and in this study it performed better than two other classifiers (NB and DT) in analysing the sentiment on selected user-generated content such as reviews or feedback about a particular entity or subject. In the testing phase, the solution scores high in term of accuracy for every dataset tested.

Then, we applied VADER on the reviews or feedbacks from student toward an e-learning in an institution. To optimise the results, we normalize several words to match to the words in VADER dictionary. The reason behind that was in some cases VADER was unable to detect either the feedback is negative or positive in some sentences which are clearly negative or positive for the reason of no sentiment word in the sentences. So, by doing normalization in pre-processing we can tackle that issue to an acceptable level.

Following are some advantages which were identified in this research:

1) This research had proven that VADER was better than two ML classifiers (NB and DT). So, it can be applied to the extract the sentiment and deal with informal short text generated by users who want to express their satisfaction or dissatisfaction to particular entity.
2) To give some exposure to other researchers on applying two types of techniques in

sentiment
3) analysis (ML and lexical-based).
4) To give some exposure to other researchers on using several tools like Python, NLTK and Text Blob in SA.

To inspire other researchers to do more research on SA using these techniques or to continue and improve this project so that a better result will be retrieved.

ACKNOWLEDGEMENT

This research was supported by Research Management and Innovation Centre, Universiti Sultan Zainal Abidin.

REFERENCES

[1] Hutto CJ, Gilbert E. "*Vader: A parsimonious rule-based model for sentiment analysis of social media text*". Eighth Int AAAI Conf Weblogs Soc Media [Internet]. 2014;216–25. Available from: http://www.aaai.org/ocs/index.php/ICWSM/ICWSM14/paper/view/8109%5Cnhttp://comp.social.gatech.edu/papers/icwsm14.vader.hutto.pdf

[2] Chaturvedi I, Cambria E, Welsch RE, Herrera F. "*Distinguishing between facts and opinions for sentiment analysis: Survey and challenges*". Inf Fusion [Internet]. 2018;44(December 2017):65–77. Available from: https://doi.org/10.1016/j.inffus.2017.12.006

[3] Dhanalakshmi V, Bino D, M SA. "*Opinion mining from student feedback data using supervised learning algorithms. MEC Int Conf Big Data Smart City*". 2016;1–5.

[4] Szerszen D, Palsson A. "*Sentiment Classification in Social Media*". KTH Royal Institute of Technology; 2016.

[5] Bhadane C, Dalal H, Doshi H. "*Sentiment analysis: Measuring opinions*". Procedia Comput Sci [Internet]. 2015;45(C):808–14. Available from: http://dx.doi.org/10.1016/j.procs.2015.03.159

[6] Devika MD, Sunitha C, Ganesh A. "*Sentiment Analysis: A Comparative Study on Different Approaches*". Procedia Comput Sci [Internet]. 2016;87:44–9. Available from: http://dx.doi.org/10.1016/j.procs.2016.05.124

[7] Medhat W, Hassan A, Korashy H. "*Sentiment analysis algorithms and applications: A survey*". Ain Shams Eng J [Internet]. 2014;5(4):1093–113. Available from: http://dx.doi.org/10.1016/j.asej.2014.04.011

[8] Kotzias D, Denil M, de Freitas N, Smyth P. From "*Group to Individual Labels Using Deep Features*". Proc 21th ACM SIGKDD Int Conf Knowl Discov Data Min [Internet]. 2015;597–606. Available from: http://doi.acm.org/10.1145/2783258.2783380

[9] Haddi E, Liu X, Shi Y. "*The role of text pre-processing in sentiment analysis*". Procedia Comput Sci [Internet]. 2013;17:26–32. Available from: http://dx.doi.org/10.1016/j.procs.2013.05.005

[10] Fayyad U, Piatetsky-Shapiro G, Smyth P. The KDD "*Process Knowledge from Volumes of Data*". 1996;39(11):27–34

[11] R. Rosly, M. Makhtar, M. K. Awang, M. N. A. Rahman, and M. M. Deris, "*Multi-classifier models to improve accuracy of water quality application*," ARPN Journal of Engineering and Applied Sciences, vol. 11, no. 5, p. 3209, 2016.

[12] M. K. Awang, M. Makhtar, M. N. A. Rahman, and M. M. Deris, "*A new soft set-based pruning algorithm for ensemble method*," Journal of Theoretical and Applied Information Technology, vol. 88, no. 3, pp. 384–391, 2016.

[13] M. Makhtar, Y. Longzhi, D. Neagu, and M. Ridley, "*Optimisation of classifier ensemble for predictive toxicology applications*," Computer Modelling and Simulation (UKSim), 2012. Available: http://www.academia.edu/4750498/Optimisation_of_Classifier_Ensemble_for_Predictive_Toxicology_Applications. [Accessed: 15-Apr-2015].