Evaluation of E-learning Activity Effectiveness in Higher Education Through Sentiment Analysis by Using Naïve Bayes Classifier

Eka Angga Laksana
Widyatama University
eka.angga@widyatama.ac.id

Ase Suryana
Widyatama University
ase.suryana@widyatama.ac.id

Ai Rosita
Widyatama University
ai.rosita@widyatama.ac.id

Heri Heryono
Widyatama University
heri.heryono@widyatama.ac.id

***Abstract* – Sentiment analysis being a part of text mining research domain has been recognized due to their successful implementation in social media analysis. Sentiment analysis methods had intelligent ability to classify texts into negative or positive. Classified texts concluded whole users respond and described opinion polarity about particularly topic. Based on this idea, this research took e-learning's users opinion as object measured through sentiment analysis. The results can be used to evaluate the e-learning activity. This research had been implemented in Widyatama University which had been running e-learning activity for several years. Qualitative method by given questioner to users and gather the feedback is commonly used as evaluation of e-learning system previously. Still, questioner doesn't represent the conclusion about the whole opinion. Hence, it needs the method to identify opinion polarity from e-learning member. The e-learning opinion data sets were gathered from questioner filled by e-learning member included both student and lecturer as participants. The participants gave review about learning outcome after their participation in e-learning activity. Their opinion was needed to describe current situation about e-learning activity. Therefore, the conclusion could be used to make improvement and describe few achievements about the e-learning system. The data sets trained by Naïve Bayes classifier was grouped into negative or positive in its user respond. The classification results were also evaluated by a number of particular evaluation metric used in data mining to show the classifier performance such as accuracy, precision, and recall.**

Keywords: classifier, evaluation, e-learning, sentiment, naïve bayes

I. INTRODUCTION

The rapid growth of the internet based technology leads to revolution in education area. The traditional face to face learning changed by web based learning and relieve distance during learning process [1]. Web based learning which known as e-learning system has numerous approach in different educational level. Now days a lot of educational institution offering e-learning as part of learning activity especially in higher education [2][1]. E-learning system defined as the improvement of teaching quality through multimedia and internet technologies. E-learning not only provide teaching material and educational service but also evaluation scheme, exchange information and collaborative learning between student and lecturers [3][4]. The main purposes behind Education institutions use the e-learning technology is to improve learning outcomes. They believe that more participation, self-regulated and interactive communication are considered as key factor to improve learning outcomes [2][5]. Recently, e-learning platforms also introduced in public administrations and corporations to make learning quality better than traditional teaching [6][7].

E-learning effectiveness contain detailed information based on evaluation result of e-learning system. It can lead to reflection and revision of the learning approach adopted in educational institutions [8][9]. Usually educational institutions do the assessment into their e-learning system to know the effectiveness which useful to know learning outcomes that has been achieved and possibly compared current or previous method and approach. E-learning effectiveness sometimes difficult to measure empirically without controlling all involved variable which still difficult to do in real environment. This paper purposes to perform the analysis of the e-learning system effectiveness based on personal feedback retrieved from both student and teacher. We believe that "what other people think" has become important piece of information for the people during decision making process [10]. Therefore, the personal impression retrieved from student and teacher could be a good indicator about e-learning effectiveness and useful for top level management to create decision.

Sentiment or opinions are key of our activity because human behavior influenced based on it. Human decision sometimes influenced by the other opinion. In the real world, organization and business always try to improve their product and service by finding out about customer's opinion. It is commonly happened because customer always want to know about another opinion of a product before purchasing it. This paper conducted based on this idea, whether the e-learning participant opinion would lead to proper advice for decision making process.

The remainder of this paper is organized as follows: Review of literature on e-learning evaluation and previous research in sentiment analysis; Discusses the research method of this study included data preparation and experimental setup; Discusses the experiment result and analysis; Finally summarize the result and suggest the possibility for the next research.

II. REVIEW OF LITERATURE

Research in e-learning becomes interesting as more and more higher education in the world wide has been used e-learning system for their course delivery and tried to comprehend how effective and usable related to interaction between human and computer [11][12][13]. Positive user experience could indicates the acceptance, satisfactions and efficiency of academic institutions [14]. The system itself is not adequate to sustain new educational approach like e-learning, therefore higher education must understand and learn whether users have got positive or negative experience during their study [15][5]. E-learning system meet the requirements when users satisfy and feel positively [16][17][18].

About Moodle

Moodle originally design by Martin Dougmias, was released on 20 August 2002. Moodle is known as robust open source e-learning platform was developed by collaboration effort of international community. Until now, Moodle e-learning platform still continually improved to give teacher, administrator and student with a stable, secure and increased learning experience. Currently, Widyatama University adapt "Blended Learning" as e-learning method approach. Blended Learning is learning approach that uses both face-to-face and e-learning [19]. Classroom based teaching enable the student to consolidate their skill and knowledge. It usually held on first and near exam meeting, while the online learning has been held at the rest schedule. Online learning by Moodle allows the student to gain the resource and doing activity to make sure that they can revise their works, discuss in forum and involve in quiz.

III. METHODS

The data was collected from e-learning system at Widyatama University. We create questionnaire form opened 5 January 2017 and closed on 4 February 2017. The questions were asked to the e-learning participant from both teacher and student through questionnaire form. The whole of Participants was recently use moodle e-learning platform as learning activity in Widyatama University. They are registered e-learning member on running semester which have been enrolled in moodle class during running semester. The participants must fill their opinion about e-learning system that they have experienced before.

The e-learning user's response corpus has 272 positive and 168 negative sentences. This research has used 4/5 of them as training set, and the rest as test set. This constructs dataset containing 351 training instances and 89 test instances. The naïve Bayes classifier training method has created a token list in the form of [(feats, label)], where feats is feature dictionary and label the classification label. Feats contains {word:True} and label will be 'positive' or 'negative'. This study, we assume the corpus as direct opinions. They are easier to handle, otherwise indirect opinions often needs more time to deal with [20]. For evaluation methods, this study uses nltk.classify.util.accuracy, nltk.precision, nltk.recall and nltk.f_measure library [21].

This research uses NLTK (Natural Language Toolkit), a python based programs which known as platform to work with human language data [21]. Naïve bayes has been used as base classifier algorithm to train corpus. The experiment which has been conducted in this research follows several steps as described below:

Collect the data through questioner. The questioner form form has been distributed to e-learning participant for both student and teacher in certain periods of time.

Data preprocessing. Retrieving the participant's feedback. Create separation manually toward dataset into positive and negative corpus.

Feature extraction. Deciding the relevant feature for classifier by selecting specific words.

Training and testing dataset. Doing cutoff for both positive and negative corpus as much as 80% for training set and 20% for testing.

Classify using naïve Bayes classifier. Implementing machine learning algorithm to learn word pattern that represent sentiment.

Performance evaluation. Performing the evaluation scheme include accuracy, precision, recall, and F-measure metric.

Conclude the sentiment result. Extracting the most important feature based on classification result. This is useful for higher education to create decision for future improvement on e-learning system.

Naïve Bayes Classifier

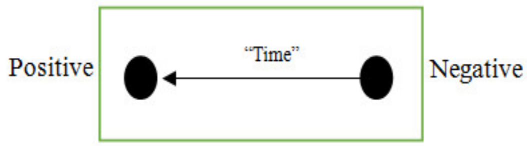

Figure 1. Classification degree

Above figure illustrate training corpus which most classified into negative so the classifier starts closer to the" negative" label. In this example, the input document contains the word "time" which strong indicator for "positive" label. After each feature has made its contribution, the naïve bayes checks which it is indicated to, and defining that label to the input. For example, the word "time" occur in 80% of the positive document, 20% in negative document. Calculated likehood score, by multiplied by 0.8 for the positive label and 0.2 for negative label. The whole effects, will be to decrease the score of the negative more than positive label.

Naïve bayes algorithm creates classification by finding the probability for a label. First, it uses the Bayes rule define Q (label|features) in term of Q (label) and Q(features|label) and N(feature|label)

$$Q(label|features) = \frac{Q(label) * Q(features|label)}{Q(features)} \quad (1)$$

Naïve Bayes algorithm then makes the 'naïve' assumption which whole features are independent as formulated below:

$$Q(label|features) = \frac{Q(label) * Q(features|label) * \dots * Q(fn|label)}{Q(features)} \quad (2)$$

Then for each label, the algorithm calculates the numerator and normalized them by sum to one as represented in the following formula:

$$Q(label|features) = \frac{Q(label) * Q(features|label) * \dots * Q(fn|label)}{sum[l](Q(l) * P(f1|l) * \dots * P(fn|l))} \quad (3)$$

Accuracy has been commonly used to evaluate a classifier. It shows the percentage of test set which is correctly labeled. In this study uses nltk.classify.accuracy(Steven et al., 2009) to calculate the accuracy classified sentiment on a given test set.

Since sometimes the number of relevant document lower that irrelevant document, the accuracy scores for irrelevant labeled document would be near to 100%. Therefore, there are four terms to represent different set of measures. Relevant items correctly identified as relevant defined as "True positives" (TP). irrelevant items which correctly identified as negative defined as "True negative" (TN). "False positive" (FP) defined as irrelevant items that are incorrectly identified as relevant, and finally "false negative" (FN) as relevant items which is incorrectly identified as irrelevant.

Precision and recall is another performance evaluation which tried to overcome shortcomings of accuracy. Because sometimes accuracy can be misleading in "search task" while attempting to find data which relevant to an appropriate task. Precision indicates the number of relevant items which identified were relevant, with the formula is *TP/(TP+FP).* Recall shows the number of relevant items which successfully identified by the formula TP/(TP+FN). F-measure (F-score) combined by precision and recall, also defined as harmonic mean of precision and recall by the formula (2 x precision x recall) / (precision + recall).

IV. RESULTS AND DISCUSSIONS

Accuracy

After the evaluation process, the algorithm shows accuracy on 87.5%. It means 87.5% was correctly labelled on the test set.

Precision, Recall and F-Measure

As shown on table 1, 93% recall means every user's responses that has identified positive correctly. Very few false negatives in the positives class. Correct positive classification identified with 87% positive precision and the rest 13% identified as false positive for positive label. Negative precision as 87.9% indicates very few false positive for the negative class. Relatively high recall causes about 23% false negative for negative label. F-Measure shows weighted harmonic means between precision and recall.

Informative Features

Table 1. Precision and recall score

pos precision:	0.872852233677
pos recall:	0.933823529412
pos F-measure:	0.902309058615
neg precision:	0.879194630872
neg recall:	0.779761904762
neg F-measure:	0.826498422713

Table 1 shows the most informative feature which affect the sentiment degree. The table only shows ten informative features sorted by their ratio in feature label. The table also shows that the training set which contain the word "kuis" reaches 13.5 labeled as negative more often than positively labeled. Another feature is the word "forum" also labeled as negative more than positive as much as 10.7 times higher.

Table 2. Most informative feature

word	Feature label	
contains(kuis)=True	negative : positive=	13.5 : 1.0
contains(forum)=True	negative : positive=	10.7 : 1.0
contains(quiz)=True	negative : positive=	10.2 : 1.0
contains(kurang)=True	negative : positive=	10.1 : 1.0
contains(sulit)=True	negative : positive=	9.2 : 1.0
contains(tetapi)=True	negative : positive=	9.2 : 1.0
contains(ketika)=True	negative : positive=	8.1 : 1.0
contains(dosen)=True	negative : positive=	7.9 : 1.0
contains(apalagi)=True	negative : positive=	7.0 : 1.0
contains(menggunakan)=True	negative : positive=	7.0 : 1.0

Table also shows the result that needed by top level management to create decision making. The words that appear in most informative feature can be analyzed to find the major drawback on current e-learning system. In addition, selecting relevant feature and deciding how to handle with it for learning process leads into learning model ability to create a good model. In this study, we use a fairly simple and obvious set of features which is carefully constructed during data preparation. Typically, feature extraction involves trial-and-error process guided by intuition about what information is correct related to the problem.

V. CONCLUSION AND SUGGESTIONS

This research performed classification on e-learning participant's opinions. The opinion datasets have been labeled into positive and negative which divided into training and testing set. Naïve Bayes algorithm has been used as learning method and shows the accuracy by 87.5%. Another evaluation also performed with precision, recall and F-measure to represent relevant and irrelevant document and they show a good result. It means the model has been successfully used to classify opinion and extract the most important features to be used by top level management to create decision making. The critical point that represent drawback and effectiveness has been shown on Table 2. More analyzed by using simple statistic can be used by utilizing each word contained in the most informative feature as the main keyword to create improvement and learning outcome achieved by student. Another improvement could be made to achieve better result since this study focus on direct opinion. As suggestion for the research, the different type of opinion method can be used to improve machine learning ability.

REFERENCE

[1]. Guri-rosenblit, S. (2005). Eight Paradoxes in the Implementation Process of E-learning in Higher Education, 5–29. https://doi.org/10.1057/palgrave.hep.8300069.

[2]. Garrison, D. R., & Anderson, T. (2011). E-Learning in the 21st Century A Framework for Research and Practice.pdf. Taylor & Francis.

[3]. Dominici, G., & Palumbo, F. (2013). How to build an e-learning product : Factors for student / customer satisfaction. *Business Horizons*, *56*(1), 87–96.

[4]. Nichols, M. (2008). Institutional perspectives: The challenges of e-learning diffusion, *39*(4), 598–609. https://doi.org/10.1111/j.1467-8535.2007.00761.x.

[5]. Persico, D., Manca, S., & Pozzi, F. (2014). Computers in Human Behavior Adapting the Technology Acceptance Model to evaluate the innovative potential of e-learning systems. *Computers in Human Behavior*, *30*, 614–622. https://doi.org/10.1016/j.chb.2013.07.045.

[6]. Stoffregen, J., Pawlowski, J. M., & Pirkkalainen, H. (2015). Computers in Human Behavior A Barrier Framework for open E-Learning in public

administrations. *COMPUTERS IN HUMAN BEHAVIOR*. https://doi.org/10.1016/j.chb.2014.12.024.

[7]. Violante, M. G., & Vezzetti, E. (2013). Virtual Interactive E-Learning Application : An Evaluation of the Student Satisfaction, 72–91. https://doi.org/10.1002/cae.21580.

[8]. Lee, K. (2005). E-Learning : The Quest for Effectiveness, *2*.

[9]. Novo-corti, I., Varela-candamio, L., & Ramil-díaz, M. (2013). Computers in Human Behavior E-learning and face to face mixed methodology : Evaluating effectiveness of e-learning and perceived satisfaction for a microeconomic course using the Moodle platform. *Computers in Human Behavior*, *29*(2), 410–415. https://doi.org/10.1016/j.chb.2012.06.006.

[10]. Pang, B., & Lee, L. (2008). Opinion Mining and Sentiment Analysis, *2*, 1–135. https://doi.org/10.1561/1500000001.

[11]. Bringula, R. P. (2013). Computers & Education In fl uence of faculty- and web portal design-related factors on web portal usability : A hierarchical regression analysis. *Computers & Education*, *68*, 187–198. https://doi.org/10.1016/j.compedu.2013.05.008.

[12]. Escobar-rodriguez, T., & Monge-lozano, P. (2012). Computers & Education The acceptance of Moodle technology by business administration students. *Computers & Education*, *58*(4), 1085–1093. https://doi.org/10.1016/j.compedu.2011.11.012.

[13]. Navimipour, N. J., & Zareie, B. (2015). Computers in Human Behavior A model for assessing the impact of e-learning systems on employees ' satisfaction, *53*, 2011–2013.

[14]. Hubackova, S. (2014). Effectiveness and evaluation of on-line courses. *Procedia - Social and Behavioral Sciences*, *143*, 139–142. https://doi.org/10.1016/j.sbspro.2014.07.375.

[15]. Laurillard, D., Wasson, B., M., & Hoppe, U. (2009). *Implementing technologyenhanced learning. In Technology-enhanced learning*.

[16]. Capece, G., & Campisi, D. (2013). User satisfaction affecting the acceptance of an e- learning platform as a mean for the development of the human capital, (April), 37–41. https://doi.org/10.1016/j.bushor.2012.09.011.

[17]. Lee, S. M., Kim, Y. R., & Lee, J. (2016). An Empirical Study of the Relationships among End-User Information Systems Acceptance , Training , and Effectiveness An Empirical Study of the Relationships among End-User Information Systems Acceptance , Training , and Effectiveness, *1222*(April). https://doi.org/10.1080/07421222.1995.11518086.

[18]. Yeh, Y., & Lin, C. F. (2015). Aptitude-Treatment Interactions during Creativity Training in E-Learning : How Meaning-Making , Self-Regulation , and Knowledge Management Influence Creativity, *18*, 119–131.

[19]. Graham, C. R. (2003). DEFINITION, CURRENT TRENDS, AND FUTURE DIRECTIONS.

[20]. Liu, B. (2012). Sentiment Analysis and Opinion Mining, (May).

[21]. Steven, B., Klein, E., & Loper, E. (2009). *Natural Language Processing with Python*.