Sentiment analysis of preschool teachers' perceptions on ICT use for young children

Mohd Ridzwan Yaakub
Center for Artificial Intelligence Technology (CAIT),
Faculty of Information Science and Technology
Universiti Kebangsaan Malaysia
Bangi, Selangor, Malaysia
ridzwanyaakub@ukm.edu.my

Filzah Zahilah Mohamed Zaki
Department of Science and Technical Education,
Faculty of Educational Studies,
Universiti Putra Malaysia
Serdang, Selangor, Malaysia
gs52106@student.upm.edu.my

Muhammad Iqbal Abu Latiffi
Center for Artificial Intelligence Technology (CAIT),
Faculty of Information Science and Technology
Universiti Kebangsaan Malaysia
Bangi, Selangor, Malaysia
p95662@siswa.ukm.edu.my

Susan Danby
School of Early Childhood and Inclusive Education,
Faculty of Education
Queensland University of Technology
Brisbane, QLD, Australia
s.danby@qut.edu.au

***Abstract*—Sentiment analysis in gaining more attention as it is increasingly used in multiple domains, including in interpreting educational data. The article uses sentiment analysis technique to understand the early childhood educators reported beliefs (perception) on young children's ICT use. The dataset was obtained from a comparative study of early childhood educators from two countries, Australia and Malaysia. The result shows a similar outcome where most teachers agreed upon the benefits of ICT use and conclude more positive sentiment polarity.**

This paper summarizes the findings using sentiment analysis as well as comparing it to the quantitative data obtained from the survey.

Keywords— educational data analytics, sentiment analysis, machine learning, qualitative data

I. INTRODUCTION

In recent years, a collaboration of interdisciplinary, or multidisciplinary nature has emerged as a new trend in research. For instance, the application of computer science principles in the education domain is gaining more popularity. Obviously, it is due to the potentials that computer science can offer in general, and the convergence of artificial intelligence in various field, including in analyzing educational data. More recently, the collaboration of interdisciplinary research is evident for both computer science and education field. There are multiple ways of how these two fields can work together in research including in Educational Data Mining (EDM), Learning Analytics and Knowledge (LAK), Computational Thinking (CT), as well as computing education itself, to name a few. Sentiment Analysis (SA) too, has great potential, yet it remains underused in educational context [1] despite its extensive use in other disciplines. Also, there is a need to foster more communication and collaboration between LAK and EDM to reap more benefit in research, by sharing of methods, and tools for data mining and analysis [2]. According to the Society of Learning Analytics Research, "Learning analytics is the measurement, collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimizing learning and the environments in which it occurs." [3]

Hence, the article aimed to re-introduce this artificial intelligence tool, by deploying sentiment analysis as another data analytical means. This technique offers an alternative to the conventional way (such as thematic analysis) to gauge qualitative data and to compare it with the related quantitative data, in this case, it was derived from a comparative study in education.

II. RELATED WORKS

A. *Teachers perceptions and adoption of ICT*

Over the years, a growing body of research in an educational setting has been discussing teachers' attitude and their uptake of ICT, including for preschool teachers. The issue is one of the major areas of interest within the field of educational technology. Most of the study has applied their preferred lenses in investigating the problems and offering different perspectives within various contexts. Many studies reported while the majority of teachers has positive beliefs, this may not be necessarily true in their classroom practice, due to various factors, including structural issues or lack of technical support [4]. This is despite policies related to ICT availability in classrooms. Other factors might contribute to this positive belief and the uptake of ICT in an educational setting including the need for continuous teachers' professional development [5], among others. The original study where the data on preschool teachers' beliefs (or perception) benefits of ICT use on young children has the following variables, as shown in Fig. 1 below:

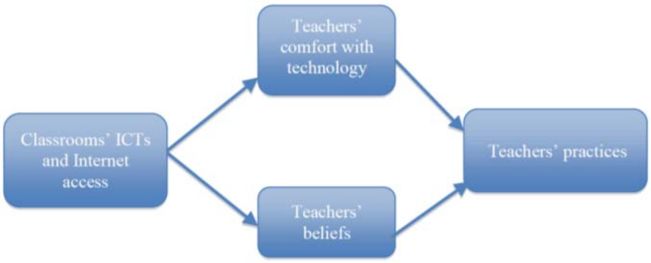

Teachers' comfort with technology
Classrooms' ICTs and Internet access
Teachers' practices
Teachers' beliefs

Fig.1 Preschool teachers' beliefs on ICT use by young children, as part of the study [6]

B. Sentiment analysis in educational research

Among examples of sentiment analysis use in educational context is sentiment analysis in student learning experience [7] and teachers feedback [8], [9] and learning outcome-based feedback [10]. For instance, [11] suggested the use of Natural Language Processing (NLP) and machine learning be applied to student feedback to assist university administrators and teachers in handling teaching and learning issues, by using students' comments from course surveys and online sources to identify sentiment polarity, the emotions expressed, and satisfaction. In the context of e-learning, [12] recommended the use of a hybrid approach to get the best accuracy.

C. Sentiment Analysis Techniques

Sentiment Analysis or Opinion Mining techniques can be separated into three different groups which are supervised, semi-supervised and unsupervised techniques. The supervised techniques are machine learning classifiers. Supervised techniques more accurate, however, need to be trained on a relevant domain [13]. Unsupervised statistical techniques do not require training. They are efficient in a dynamic environment but at the cost of accuracy. Sentiment analysis techniques analyze opinion datasets to generate a general perception that people have about products, services or thoughts on a certain issue. The classification of sentiments in a review document is performed by identifying and separating all the positive and negative opinion words. Considering the strength of these words, along with their polarity, helps in multi-class classification [13], [14]. Machine learning techniques suitable with sentiment analysis task as plenty of data and its presence in pattern [13]. A test set dataset is used to figure out the performance of the classifiers. These classifiers can also be successful for classification of text effectively [15]. The classifiers are trained on a labeled dataset which called supervised learning. The supervised learning approach relies on the existence of labeled training dataset [16].

Naïve Bayes (NB) classifier is popular for text classification. NB can determine the possibilities of the classes to relate the review with the help of feature vector [13], [14]. The feature vector actually labeled as positive and negative for binary classification. By using the annotated training data set the probability of feature vector can be calculated with each label. A label is assigned by the feature vector that has the highest probability for it [13], [16]. If this information is sustained, it can be used to show the confidence in a label for a feature vector. In [17] they used Naïve Bayes classifier in their study to classify movie reviews. From their study, the found that NB classifier produced outperformance where 0.91 was their accuracy. Besides that, Naïve Bayes classifier also performed well in movie reviews where the accuracy was 81% [18]. For reviews from microblog that done by [10], the accuracy for Naïve Bayes was 83%. NB has succeeded in producing better performance in several experiments that involving different type of domains.

Support Vector Machine (SVM) is one of the popular methods for text classification. Also, SVM one of the most efficient approaches for text classification as found in many studies [19]–[21]. SVM is effective in the text analysis that works by developing a hyper-plane to two separate classes which are positive and negative while maximizing the margin between the two classes [21]. Support vectors calculate the margin that is formed, one on each side of the hyper-plane [13], [21], [22]. The massive amount of time that needed, which is related to the number of training instances and found unreliable for large scale application is one of the limitations for SVM [13], [16], [21].

The k-nearest neighbour is a type of instance-based earning or non-generalizing learning where it does not need to construct a general internal model but simply stores instances of the training data [17]. When the irrelevant features were removed, the accuracy of the model improves. Features are assigned weights to vary their contribution towards decision making [13]. Weights are extracted from the probability of information in documents across different categories. A study on performance of k-NN using a pre-processed dataset that mentioned in [13] claiming 10 % improvement when noise and outliers are moved out. Optimum value is chosen as a threshold to separate regular data from noise. Sentiment analysis is performed with a reduced set of the feature vector in [23] to avoid the disaster of dimensionality. It is effective against large feature sets and the order of features and their thresholds are identified from within the training data [13]. k-NN has encouraging results in sentiment analysis. But, it is more susceptible to noise and high dimensional feature set. Therefore, more of the work in k-NN for text classification has focused on feature selection and reduction techniques as they are the driving factors of k-NN's performance [13].

III. Material and Method

A. The educational study

The datasets were taken from a comparative study investigating preschool teachers' comfort levels, and practices and beliefs of young children's ICTs and Internet use in early childhood contexts in both Australia and Malaysia. [6]. The study used a survey questionnaire as the instrument to collect data by using both online and printed copies to reach target samples in both countries. The study involved a total of N=131 for Australian participants and N=103 for Malaysian counterparts.

B. Dataset

For this study, we analyzed the qualitative data from an educational research. The survey questionnaires have a part where the teachers express their thought about children experience through accessing the Internet and Web searching. One of the survey questions in this study: "What do you believe children experience through accessing the Internet and Web searching?" was an open-ended question at the end of the questionnaire with opinions from teachers where mentioned in [24] text comments one of the unstructured data that will be analyzed using sentiment analysis tool in this article. The survey data has both positive, negative and neutral reviews. The sentiment for this dataset has been annotated by using Valence Aware Dictionary for sEntiment Reasoning (VADER), a lexicon-based method of classifying reviews or comments to a particular sentiment. The method of classifying the sentiment using VADER is also often used in other research works [25]–[27].

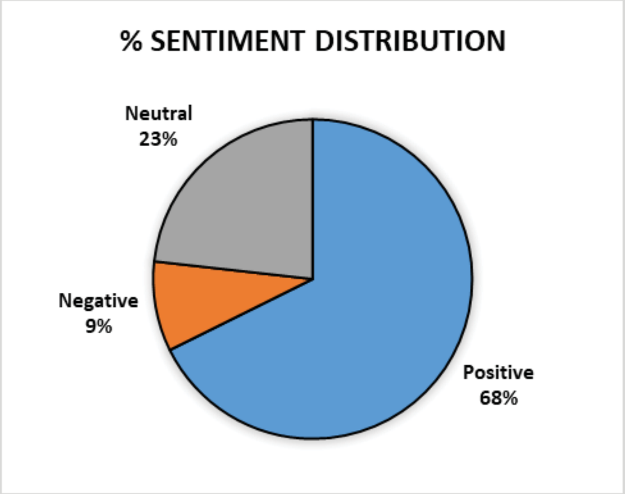


Fig. 2. Percentage sentiment distribution

C. Methodology

This study specifically constructed to understand early childhood teachers' current educational beliefs of their general ICT and Internet use in the preschool years of education in Malaysian and Australian context.

The data was in CSV file format. The data then was converted into arff format. The arff file was imported into the WEKA data mining tool. WEKA is an abbreviation from Waikato Environment for Knowledge Analysis where this tool was developed by University of Waikato, New Zealand. This software can provide an implementation of machine learning algorithms and contains a module for data processing. In addition, it supports several data mining tasks includes data preprocessing, binning, feature selection, clustering and regression.

The main components of the methodology consist of 4 phases as shown in Fig. 3:

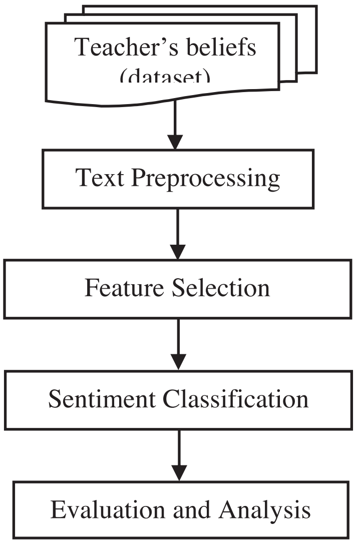


Fig.3 Process of Sentiment Analysis

Phase 1: Text Preprocessing is an essential part of any Natural language Programming (NLP) system since the characters, words and sentences identified at this stage. These are the fundamental units passed to all further processing stages, from analysis and tagging components [28]. First, the researchers need to eliminate noisy data from the raw dataset before proceeding to the next phases where tokenization took place. Tokenization is the process of breaking a stream of text into words, phrases, symbols or other meaningful elements called tokens [29]. The list of tokens becomes input for further processing such as parsing or text mining. Textual data is only a block of characters at the beginning. All processes in information retrieval require the words of the data set. So the requirement for a parser is tokenization of documents. The main use of tokenization is identifying meaningful keywords

Then normalization occurred where the process involves transforming text to ensure consistency. Some examples of this process include converting upper case letter to lower case, Uni-code conversation and removing diacritics from letters, punctuations or numbers [29]. Stemming is one of the important roles in NLP where the process of conflating the variant forms of a word into a common representation, which is the stem. For example, the word "presentation", "presented", "presenting" could all be reduced to a common presentation "present" [29], [30].

Next, researchers performed stop word removal where many words in documents recur very frequently but are essentially meaningless as they are used to join words together in a sentence. It is commonly understood that stop words do not contribute to the context or content of textual documents. Due to their high frequency of occurrence, their presence in text mining presents an obstacle in understanding the content of the documents. Stop words are very frequently used common words such as 'and', 'are', 'this' and etc. These words are not useful in the classification of documents. That's the reason it must be removed [29], [31].

Phase 2: Classification performances can be improved when feature selection is applied where it will remove irrelevant and redundant features from the actual dataset [32]. Attribute selection filter or feature selection is applied to decrease the dimensionality of the dataset and can improve the learning algorithms' performances. There are three techniques in feature selection: filter, wrapper and hybrid [33]. In this study, Correlation Attribute Eval (CA) where it evaluates the attributes with respect to the target class [34] is evaluated.

Phase 3: There were three supervised learning algorithms were used to perform the classification task. The algorithms are Naïve Bayes, Support Vector Machine (SVM) and Random Forest. The measurement that will be considered is the percentage of correctly classified instances and for the validation part, the researcher used the 10-fold cross validation.

Phase 4: Evaluation and analysis: The verification is conducted by comparing the result the acquired from the reviews that already classified from different classifiers. The result will be discussed in the next part of this paper.

IV. Experiment and Result

This study is completely experimental with a practical implementation. Data visualization can be utilized to get an overview of the pattern of the dataset in a graphical view. Fig. 4 shows visualization for attribute that consists of three different sentiment distribution where positive sentiment represents by blue, negative sentiment represented by light blue and neutral sentiment represent by red.

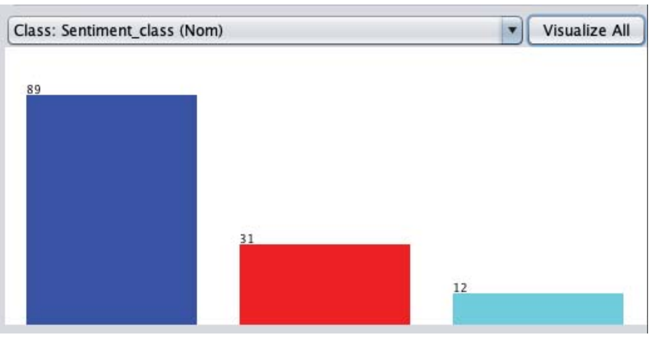
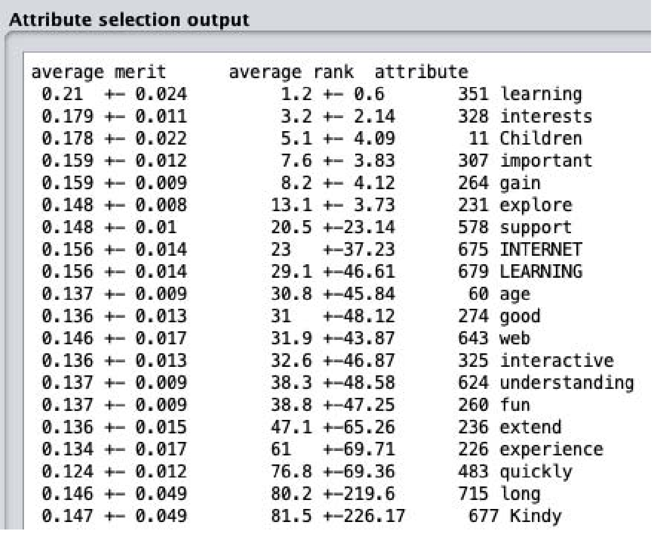


Fig. 4. Data visualization for attribute sentiment_class.

The experiment of feature selection on the dataset is conducted using Correlation Attribute. Fig. 5 shows the top 20 attributes for the ranking method after went through feature selection process.



Attribute selection output

```
average merit      average rank  attribute
 0.21  +- 0.024         1.2 +- 0.6       351 learning
 0.179 +- 0.011         3.2 +- 2.14      328 interests
 0.178 +- 0.022         5.1 +- 4.09       11 Children
 0.159 +- 0.012         7.6 +- 3.83      307 important
 0.159 +- 0.009         8.2 +- 4.12      264 gain
 0.148 +- 0.008        13.1 +- 3.73      231 explore
 0.148 +- 0.01         20.5 +-23.14      578 support
 0.156 +- 0.014        23   +-37.23      675 INTERNET
 0.156 +- 0.014        29.1 +-46.61      679 LEARNING
 0.137 +- 0.009        30.8 +-45.84       60 age
 0.136 +- 0.013        31   +-48.12      274 good
 0.146 +- 0.017        31.9 +-43.87      643 web
 0.136 +- 0.013        32.6 +-46.87      325 interactive
 0.137 +- 0.009        38.3 +-48.58      624 understanding
 0.137 +- 0.009        38.8 +-47.25      260 fun
 0.136 +- 0.015        47.1 +-65.26      236 extend
 0.134 +- 0.017        61   +-69.71      226 experience
 0.124 +- 0.012        76.8 +-69.36      483 quickly
 0.146 +- 0.049        80.2 +-219.6      715 long
 0.147 +- 0.049        81.5 +-226.17      677 Kindy
```

Fig. 5. Attribute selection output based on ranking method

After the feature selection process, the performances among the three machine learning algorithms which are SVM, Naïve Bayes and Random Forest is compared. In WEKA there has four test options available: Training Dataset, Percentage Split Supplied Test and Cross Validation. This study is conducted by using 10-Fold Cross Validation. The experiment is carried out using the 10-Fold Cross-Validation test option. Cross-validation is a statistical technique to evaluate the predictive model by splitting the original sample into two part which is training set to learn or train a model and a test set to perform the evaluation on it. For k-fold cross-validation, data are partitioned into k equally sized folds. Training and validations are performed repeatedly for each number of k-iteration.

Specifically, within each iteration, different folds of the data are also kept out for validation, while the remaining k-1 folds are retained for learning. k samples of the performance metric will then be available for each model to be evaluated. Aggregation measures such as averaging can be further performed to highlight model performance comparison, otherwise the samples can be used to support a statistical hypothesis test. Table 1 shows the comparison of the learning algorithms; Naïve Bayes, SVM and Random Forest performances in terms of recall. It is observed that the SVM algorithm outperforms other algorithms where its recall value is at 0.674 using the 10-fold Cross-Validation method.

TABLE I. PERFORMANCE OF THE ALGORITHMS ON THE DATASET

	Algorithms		
	Support Vector Machine	***Naïve Bayes***	***Random Forest***
Recall	0.674	0.606	0.659

V. DISCUSSION

This study provides a comparison of performances between three classifiers: Naïve Bayes, SVM and Random Forest in classifying sentiment using this dataset. The accuracy of these three classifiers are compared to each other. The SVM classifier achieved 0.674 recall value followed by Random Forest classifier where it achieved 0.659 recall value and recall value that achieved by Naïve Bayes is 0.606. Prior to performance comparison, several preprocessing techniques such as data cleaning and feature selection were first conducted, and the results can be increased when others learning algorithms or new hybrid algorithms is tested as well using 10-fold cross validation. The best result is given by SVM, as it works well with unstructured data such as text. Besides that, SVM consist of kernel that will give the model better strength and with a suitable kernel function, SVM can solve any complex problem.

Finding the sentiment of the reviews or comments can help in a various domain where to develop intelligent applications that can provide the users comprehensive reviews of products, movies or services. In this paper, we are reviewing the sentiments of early childhood education (or preschool) teachers on their perceptions of ICT use for young children. We also compare the result of using sentiment analysis with the related quantitative data on preschool teachers' beliefs on use of ICTs for young children (refer Fig. 6 and data in Fig. 7). From the table, there were very positive beliefs amongst the preschool teachers in both countries for the statement "it is good to use technology to build on the interest children bring to the classroom" (more than 90%), although there were also concerns on the time children spent with technology.

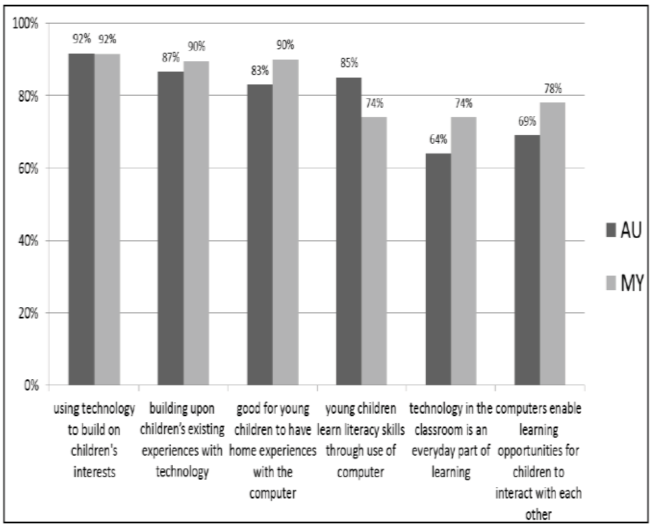
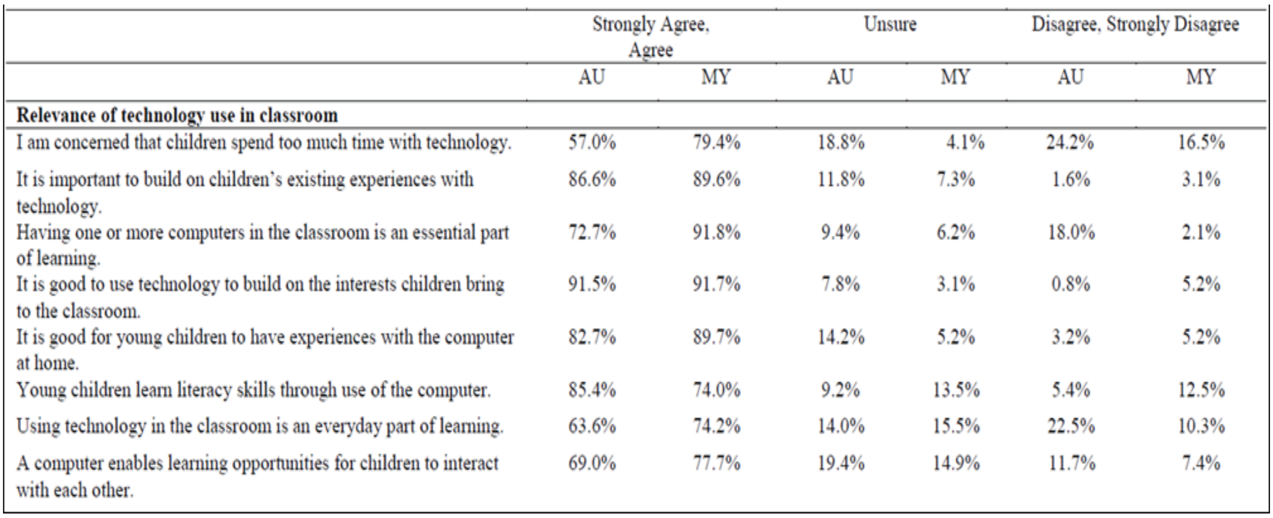


Fig. 6. Percentage of Australian and Malaysian teachers' positive beliefs on young children's technology use [6]



	Strongly Agree, Agree		Unsure		Disagree, Strongly Disagree	
	AU	MY	AU	MY	AU	MY
Relevance of technology use in classroom						
I am concerned that children spend too much time with technology.	57.0%	79.4%	18.8%	4.1%	24.2%	16.5%
It is important to build on children's existing experiences with technology.	86.6%	89.6%	11.8%	7.3%	1.6%	3.1%
Having one or more computers in the classroom is an essential part of learning.	72.7%	91.8%	9.4%	6.2%	18.0%	2.1%
It is good to use technology to build on the interests children bring to the classroom.	91.5%	91.7%	7.8%	3.1%	0.8%	5.2%
It is good for young children to have experiences with the computer at home.	82.7%	89.7%	14.2%	5.2%	3.2%	5.2%
Young children learn literacy skills through use of the computer.	85.4%	74.0%	9.2%	13.5%	5.4%	12.5%
Using technology in the classroom is an everyday part of learning.	63.6%	74.2%	14.0%	15.5%	22.5%	10.3%
A computer enables learning opportunities for children to interact with each other.	69.0%	77.7%	19.4%	14.9%	11.7%	7.4%

Fig. 7. Preschool teachers' beliefs about technology use for young children [Excerpt from [6]

In summary, there are similarities where both qualitative and quantitative data (Fig. 6 and Fig. 7) reported medium to high percentage of positive beliefs (perception) on young children use of technology. Both groups of teachers reported having positive attitudes or perceptions towards technology and Internet use in early childhood education, although these beliefs did not translate into their reported practices [6].

VI. Conclusion

The study has potential to further use sentiment analysis as a useful analytical tool within educational data mining (EDM) and Learning Analytics and Knowledge (LAK) corpus of research. Overall, this effort would contribute to the existing body of knowledge within this interdisciplinary research between education and computer science domain. The use of sentiment analysis must be extended to reach more researchers and be applicable in wider educational research objectives, has merits for researchers in both fields.

Acknowledgment

The authors gratefully acknowledge the Regional Cluster for Research and Publication (RCRP), grant number RRCP-2016-002. as well as FRGS/1/2017/ICT02/UKM/02/4 grant scheme for supporting this research.

References

[1] M. J. Koehler, S. Greenhalgh, and A. Zellner, "Potential Applications of Sentiment Analysis in Educational Research and Practice – Is SITE the Friendliest Conference ? Purpose and Research Questions," in *Society for Information Technology & Teacher Education International Conference*, 2015, pp. 1348–1354.

[2] G. Siemens and R. S. J. Baker, "Learning Analytics and Educational Data Mining : Towards Communication and Collaboration," in *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge*, 2012, pp. 252–254.

[3] D. Gasevic, "Handbook of Learning Analytics," no. April 2017, 2017.

[4] H. Mirzajani, R. Mahmud, A. Fauzi Mohd Ayub, and S. L. Wong, "Teachers' acceptance of ICT and its integration in the classroom," *Qual. Assur. Educ.*, vol. 24, no. 1, pp. 26–40, 2016.

[5] A. van den Beemt and I. Diepstraten, "Teacher perspectives on ICT: A learning ecology approach," *Comput. Educ.*, vol. 92–93, pp. 161–170, 2016.

[6] F. Z. Mohamed Zaki, "ICT and internet usage in early childhood education : a comparative study of Australian and Malaysian teachers' beliefs and current practices," Queensland University of Technology, 2013.

[7] Q. Rajput, S. Haider, and S. Ghani, "Lexicon-Based Sentiment Analysis of Teachers ' Evaluation," vol. 2016, 2016.

[8] N. Altrabsheh, M. M. Gaber, and M. Cocea, "SA-E : Sentiment Analysis for Education," in *5th KES International Conference on Intelligent Decision Technologies*, 2013, no. January 2014.

[9] P. Kaewyong, A. Sukprasert, N. Salim, and F. A. Phang, "THE POSSIBILITY OF STUDENTS ' COMMENTS AUTOMATIC INTERPRET USING LEXICON BASED SENTIMENT ANALYSIS TO TEACHER EVALUATION," in *Proceeding of the 3rd International Conference on Artificial Intelligence and Computer Science (AICS2015)*, 2015, no. October 2015, pp. 179–189.

[10] L. Huang, H. Zhao, K. Yang, Y. Liu, and Z. Xiao, "Learning Outcomes-Oriented Feedback-Response Pedagogy in Computer System Course," in *2018 13th International Conference on Computer Science & Education (ICCSE)*, 2018, pp. 1–4.

[11] S. Rani and P. Kumar, "A Sentiment Analysis System to Improve Teaching and Learning," *Computer (Long. Beach. Calif).*, vol. 50, no. 5, pp. 36–43, 2017.

[12] A. Ortigosa, J. M. Martín, and R. M. Carro, "Sentiment analysis in Facebook and its application to e-learning," *Comput. Human Behav.*, vol. 31, pp. 527–541, 2014.

[13] M. T. Khan, M. Durrani, A. Ali, I. Inayat, S. Khalid, and K. H. Khan, "Sentiment analysis and the complex natural language," *Complex Adapt. Syst. Model.*, vol. 4, no. 1, 2016.

[14] A. Al-Saffar, S. Awang, H. Tao, N. Omar, W. Al-Saiagh, and M. Al-bared, "Malay sentiment analysis based on combined classification approaches and Senti-lexicon algorithm," *PLoS One*, vol. 13, no. 4, pp. 1–18, 2018.

[15] K. Saranya and S. Jayanthy, "Learning Techniques," in *International Conference on Innovations in information Embedded and Communication Systems (ICIIECS)*, 2017, pp. 1–5.

[16] W. Medhat, A. Hassan, and H. Korashy, "Sentiment analysis algorithms and applications: A survey," *Ain Shams Eng. J.*, vol. 5, no. 4, pp. 1093–1113, 2014.

[17] M. Al Achhab and M. Lazaar, "Comparison of Feature Selection Methods for Sentiment Analysis," *Big Data, Cloud Appl.*, vol. 872, pp. 261–272, 2018.

[18] T. Sajana, C. M. Sheela Rani, and K. V. Narayana, "A survey on clustering techniques for big data mining," *Indian J. Sci. Technol.*, vol. 9, no. 3, pp. 1–12, 2016.

[19] A. Tripathy, A. Agrawal, and S. K. Rath, "Classification of Sentimental Reviews Using Machine Learning Techniques," *Procedia Comput. Sci.*, vol. 57, pp. 821–829, 2015.

[20] S. Foroozan, M. A. Azmi Murad, N. M. Sharef, and A. R. Abdul

Latiff, "Improving sentiment classification accuracy of financial news using N-Gram approach and feature weighting methods," *2015 IEEE 2nd Int. Conf. InformationScience Secur. ICISS 2015*, pp. 31–34, 2016.

[21] S. K. Trivedi and S. Dey, "Analysing user sentiment of Indian movie reviews," *Electron. Libr.*, vol. 36, no. 4, pp. 590–606, 2018.

[22] F. P. Shah and V. Patel, "A review on feature selection and feature extraction for text classification," *Proc. 2016 IEEE Int. Conf. Wirel. Commun. Signal Process. Networking, WiSPNET 2016*, pp. 2264–2268, 2016.

[23] J. Sreemathy and P. S. Balamurugan, "An Efficient Text Classification Using KNN and Naive Bayesian," *Int. J. Comput. Sci. Eng.*, vol. 4, no. 03, pp. 392–396, 2012.

[24] M. R. Yaakub, Y. Li, and Y. Feng, "Integration of Opinion into Customer Analysis Model," in *Eight IEEE International Conference on e-Business Engineering*, 2011, pp. 90–95.

[25] J. Garay, R. Yap, and M. J. Sabellano, "An analysis on the insights of the anti-vaccine movement from social media posts using k-means clustering algorithm and VADER sentiment analyzer," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 482, no. 1, p. 012043, Mar. 2019.

[26] R. Fan *et al.*, "The minute-scale dynamics of online emotions reveal the effects of affect labeling," *Nat. Hum. Behav.*, vol. 3, no. 1, pp. 92–100, Jan. 2019.

[27] A. Amin, I. Hossain, A. Akther, and K. M. Alam, "Bengali VADER: A Sentiment Analysis Approach Using Modified VADER," in *2019 International Conference on Electrical, Computer and Communication Engineering (ECCE)*, 2019, pp. 1–6.

[28] Siti Rohaidah Ahmad, Azuraliza Abu Bakar, and Mohd Ridzwan Yaakub, "Metaheuristic Algorithms for Feature Selection in Sentiment Analysis," *Sci. Inf. Conf. 2015*, pp. 222–226, 2015.

[29] S. Vijayarani, J. Ilamathi, and M. Nithya, "Preprocessing Techniques for Text Mining - An Overview," *Int. J. Comput. Sci. Commun. Networks*, vol. 5, no. 1, pp. 7–16, 2015.

[30] A. Oussous, A. A. Lahcen, and S. Belfkih, "Impact of Text Pre-processing and Ensemble Learning on Arabic Sentiment Analysis," *Assoc. Comput. Mach.*, pp. 1–9, 2019.

[31] C. S. P. Kumar and L. D. D. Babu, *Novel Text Preprocessing Framework for Sentiment Analysis*, Smart Inte., vol. 105. Springer Singapore, 2019.

[32] N. S. Sani, M. Abdul Rahman, A. Abu Bakar, S. Sahran, and H. Mohd Sarim, "Machine Learning Approach for Bottom 40 Percent Households (B40) Poverty Classification," *Int. J. Adv. Sci. Eng. Inf. Technol.*, vol. 8, no. 4–2, p. 1698, 2018.

[33] S. R. Ahmad, A. A. Bakar, and M. R. Yaakub, "A review of feature selection techniques in sentiment analysis," *Intell. Data Anal.*, vol. 23, no. 1, pp. 159–189, 2019.

[34] S. Gnanambal, M. Thangaraj, Meenatchi V T, and V. Gayathri, "Classification Algorithms with Attribute Selection: an evaluation study using WEKA," *Int. J. Adv. Netw. Appl.* , vol. 09, no. 06, pp. 3640–3644, 2018.