Bilingual Lexicon Approach to English-Filipino Sentiment Analysis of Teaching Performance

Caren Pacol and Thelma Palaoag

1Department of Information Technology, Pangasinan State University, Urdaneta City, 2428, Philippines
2Department of Computer Science, University of the Cordilleras, Baguio City, 2600, Philippines

E-mail: 1ambat_caren@yahoo.com

Abstract. The aim of this study is to formulate a strategy that can possibly calculate teacher performance by analyzing textual feedback. Expressing textual responses in quantitative form like average sentiment rating can actually provide opportunities for administrators to see if the numerical ratings given complement that of the comments. Our approach was designed to enable processing bilingual textual data. Findings of this study shows that there is strong correlation between teaching performance actual mean rating and average sentiment rating. Furthermore, the approach employed obtained 86% accuracy indicating that it is an encouraging technique, capable of analyzing the students’ textual responses. In future work, the use of POS tagging can be explored to improve sentiment analysis accuracy. Employing machine learning methods may also be considered to discover techniques and alternative approaches to sentiment classification.

1. Introduction

Teaching performance assessment is important to ensure student learning and provide basis for administrative decision making. A common practice to find out the level of teaching competency is the assessment by students. Assessment is conventionally done using survey questionnaire in each course and is composed of queries about learning undertakings.

The assessment is usually conducted by letting the student answer pool of questions using a Likert scale. The aggregated numerical ratings are used as basis to determine the teaching performance of the concerned faculty member [1]. However, assessment forms include sections for comments allowing students to freely write their observations about their teachers. The observations written on this section are usually not processed in the performance assessment due to the absence of automatic text analyzers [1, 2]. These textual feedbacks could possibly hold meaningful perceptions regarding the course, teacher proficiency, punctuality and classroom management. When analyzed effectively, these data can help administrators and the teacher reflect on what should be done to address students’ issues on their teachers. This is something that is difficult to quickly derive from the Likert scale-based feedback.

Metric based results of teaching performance evaluation are highly used to represent the quality of teaching, but it is difficult to draw interpretations from them. Though text feedbacks usually do not have weight in the calculation of teaching performance rating, they can make it easier for the administrator to identify teachers’ strengths and weaknesses. These insights can be utilized to plan for faculty development program through trainings and seminars or perform appropriate intermediary action such as coaching and mentoring that is necessary to address crucial issues in teaching performance. Since...
textual responses are more informative than numerical ratings, then expressing them in quantitative form like average sentiment rating can actually provide opportunities for administrators to see if the numerical ratings given complement that of the comments. Hence, our aim is to formulate a strategy that can possibly calculate teacher performance rating based on textual responses. This is where natural language processing come into play. Existing data mining technologies with text analytics tools can be used for natural language processing. One of the applications of natural language processing is sentiment analysis.

A person’s viewpoint conveying thoughts, attitudes and emotion is referred to as sentiment. Analyzing a group of words purposively to grasp its impression is the main focus of sentiment analysis. Sentiments are represented as either positive or negative, referred to as polarity. Looking at the sign of the polarity score, the overall sentiment is concluded as positive if it has plus (+) sign, negative if it has minus (-) sign. Students’ textual feedback in the student-faculty evaluation can be categorized as positive, negative or neutral. Thus, sentiment analysis is the appropriate approach to use in this study.

Lexicon-based sentiment analysis extract all sentiment terms for any given text and assign their sentiment value using sentiment lexicon [3]. Development of a broader scope lexicon and efficient handling of context-sensitive terms, negations and intensifiers can lead to an effective sentiment analyzer. [4]. Many sentiment analyzers that have been created were designed to evaluate single language sentences. Nevertheless, not in all cases the documents that need to be processed are expressed in single language only, especially in areas where people are multilingual. It is a common practice in sentiment analysis that sentences expressed in a different language are translated. However, translation requires maintaining the exact meaning of a sentence after translation, which is a difficult task for an automatic translator and even more difficult and time consuming when done manually. Relevant to this, we devised a lexicon of English and Filipino terms that is appropriate in the context of evaluating teaching performance in a bilingual setting. The basis of this is the similarity between the two languages such as both are written from left to right and they have similar parts of speech. This lexicon can be used to analyze students’ comments based on polarity and sentiment scores.

Analyzing students’ feedback through sentiment scoring allows measuring how intense a sentiment is instead of classifying based on positive or negative labels only. In addition, this can efficiently handle combination of positive and negative phrases in sentences. The sentiment score could either have plus (+) or minus (-) sign depending on which of the positive or negative terms has greater weight.

Specific objectives of this study are: (a) automatically calculate the mean sentiment score per faculty member (b) determine if there is correlation between the mean sentiment score in a dataset (pertaining to the comments given on one faculty member) and the actual mean rating, and (c) validate sentiment scores.

2. Related Literature

To further establish the relevance of our study, we decided to present some of lexicon-based sentiment analysis research conducted that we have reviewed.

Haider, Ghani and Rajput [1] studied on a sentiment analysis approach identifying comments with positive, negative or neutral polarity. A dictionary of sentiment terms and their labels was used. An accuracy of 91.2% was achieved in their approach. Analyzing their calculations, they found that there is correlation between sentiment scores and likert scale-based scores. Similar to [1], we used lexicon-based approach in our study. However, the difference of our strategy is that we designed our approach to process bilingual sentences incorporating valence shifters in sentiment analysis. In another study, Aung [5] designed a lexicon-based sentiment analysis framework predicting the proficiency of the teacher based on students’ text feedback. Balahadia, Juanatas and Fernando [6] also utilized a database of words in sentiment analysis in teacher evaluation. In the study by Rajput et.al, there were few misclassified feedbacks which can probably be attributed to the presence of negations in the sentences. In [5] and [6], their study included checking of negation wherein certain amount is added to or deducted from the sentiment score. Both [5] and [6] handled negations in their approaches but did not mention handling of amplifiers, de-amplifiers and adversative conjunctions. In our study, we used function that
automatically calculates sentiment scoring based on a modified lexicon of sentiment words and valence shifters.

Shaikh, Meghji and Taj [7], worked on news articles and demonstrated a lexicon-based sentiment analysis approach. When news articles have a score of +1, they were classified as positive while those yielding a score of -1 were considered negative. News articles having score of 0 were classified as neutral. An operator named “extract sentiment” was utilized to compute the sentiment rating of a full news article. This operator utilizes a SentiWordNet 3.0.0 lexicon extension which is a wordNet lexicon extension. Moreover, calculation of total sentiment rating of news article was generated using Score sentiment function.

In the paper written by Fazal Masud Kundi et.al [8], they presented a lexicon-based sentiment classification framework distinguishing and scoring slangs found in tweets. Polarity of tweets were identified as neutral, positive or negative. Comparing its results with current systems, their framework achieved higher accuracy of 92% in two-class cases and 87% in multi-class cases.

Reddy et.al [9] used the available Telugu SentiWordNet to perform sentiment analysis for Telugu eNewspapers sentences. They evaluated their proposed approach on sentiment analysis and obtained 74% accuracy on subjectivity classification and 81% on sentiment classification in news data.

The similarity of our study with Taj et.al [7], Kundi et.al [8] and Reddy et.al [9] is that it presents a lexicon-based classifier of sentences to neutral, positive or negative. However, instead of just classifying through labels, it calculates sentiment scores and aggregates them to produce average sentiment ratings. Unlike their dataset which are news articles and tweets, the dataset in this study is textual feedback of students on teaching performance. Moreover, it utilized a customized lexicon consisting of English and Filipino terms frequently used in the context of teaching performance assessment.

In [10], a method of determining polarity and calculation for lexicon-based sentiment analysis was devised and evaluated on Turkish tweets utilizing a Twitter API. An accuracy of 88.2% was achieved. To find the correct meaning of each tweet, they analyzed the data in the following levels: first is by word, second is by word group and third is by idiom/proverb.

The study in [10] employed lexicon-based approach on purely Turkish tweets. On the other hand, our study applied lexicon-based approach on English, Filipino and mixed English-Filipino sentences.

3. Methodology

The methodology is divided into the following steps.

3.1. Data Gathering

Teaching performance data were sourced from nine (9) campuses of Pangasinan State University (PSU). The data were taken from summary results of faculty teaching performance evaluation questionnaires of 367 faculty members. At present, PSU utilizes a Faculty Teaching Performance Evaluation composed of four (4) areas of evaluation namely commitment, knowledge of subject matter, teaching for independent learning and management of learning. Students rate their faculty from 1 to 5 (5 is the maximum and 1 is the minimum) based on main criteria in each area. A section for narrative comments labelled Strengths and Weaknesses is in the latter portion of the evaluation form. The results are consolidated calculating ratings in each area and the overall mean rating of each faculty member. Comments on strengths and weaknesses are also listed for individual teachers.

3.2. Data Preparation

After data were gathered, we summarized the average rating and comments corresponding to every faculty ID. Faculty members’ data were randomly selected based on the teaching performance actual mean rating. Then, comments were segregated by faculty ID to generate dataset for our experiment. The data were manually cleaned removing sentences with no clear meaning and correcting misspelled words.

Some comments were written in pure English while others were written in pure Filipino. There were also sentences that were written in mixed English-Filipino. Sample comments of students from the dataset were shown in Table 1.
The function utilized in the experiments uses a generic English-language based sentiment lexicon. Thus, we modified the sentiment lexicon to suit the context of evaluating teaching performance. More English sentiment terms had to be added and sentiment scores were modified to suit the setting. For instance, the words ‘soft’ and ‘fast’ have positive sentiment scores in the generic lexicon but have negative connotation in the context of teaching performance. We initially devised a separate Filipino sentiment lexicon using a generic dictionary. Then we added more Filipino terms that commonly occurred in students’ comments in the dataset. The function we used performs sentiment analysis at the sentence level, hence, we combined the Filipino and English sentiment lexicons. This enabled the function to analyze even the sentences written in pure Filipino as well as those written in mixed English-Filipino. The valence shifters lexicon was also modified to include Filipino terms.

Basic sentiment analysis algorithms calculate a final aggregated score by counting how many times positive and negative words occur and getting their weight or label if they occur in the sentiment dictionary. This method leads to numerous false positives when negations are not handled properly. In addition, adversative conjunctions, amplifiers and de-amplifiers have significant impact in the final sentiment score but are not accounted accordingly. Our approach simplified the procedure by using pre-defined functions that integrate English and Filipino valence shifters within a few lines of code.

To prepare our dataset for the validation of the sentiment scores, we labelled the data with positive (1), negative (-1) and neutral (0). Also, we asked two teachers to label the sentences separately. Then, we finalized the labels based on majority rule.

### Table 1. Samples of Students’ Comments.

| Sample students’ comments | Language used |
|---------------------------|---------------|
| 1 He helps the students to understand the lessons. | English |
| 2 He explain but very few explanations. | English |
| 3 Lack of patience sometimes and very strict in our quizzes and exams. | English |
| 4 Magaling magturo, maunawain. | Filipino |
| 5 Magaling siya sa explanation and also she has a good attitude. | Mixed |
| 6 Nagtuto ng mabuti and explains the lesson. | Mixed |
| 7 If the student is not learning the topic, medyo nag-iinit yung ulo niya. | Mixed |
| 8 Lagi siyang absent at kapag pumapasok hindi siya dumadating sa tamang oras. | Mixed |
| 9 Masyadong maraming requirements na ibinibigay sa amin. | Mixed |
| 10 Binibigay niya ang best niya matutunan talaga namin iyong topic na tinututuro niya. | Mixed |

### 3.3. Sentence Level Sentiment Analysis Using Sentimentr

R was used for the sentiment analysis process. First, `sentimentr` package was loaded in R. Then, data were extracted from csv file and loaded into R. The `get_sentences()` function partitions the data to individual sentences.

Throughout our experiments, we have seen that there were terms that were not listed in the lexicon which caused misclassification of the sentences. Thus, these terms were included.

Next, we utilized the `sentiment_by()` function to get mean sentiment score for a given sentence. By default, sentiment terms and valence shifters lexicons that are pre-defined in `sentimentr` package are the parameters passed to `sentiment_by()` function. A challenge we faced in the experiment is enabling the function to accept parameters that are sourced from outside the package. Both the customized sentiment terms and valence shifters lexicons were saved as csv files. A data frame with four attributes is returned by the function. An attribute labelled `ave_sentiment` is most interesting as this represents the sentiment of the sentences in one row. The resulting value could be negative or positive expressing the intensity...
and the polarity of the sentiment. In some instance the result can be 0 if the sentence contained words not found in the lexicons.

Sentiment scores are calculated by searching and comparing the words in each sentence to a lexicon of polarized words. Positive \((w_{i,j,k}^+\) words are marked with a +1 and negative \((w_{i,j,k}^-\) words, with a −1 (or other positive/negative weighting).

Words that appear four (4) words prior to and two (2) words following polarized words are considered valence shifters. When polarized word \((pw)\) is found, words nearby it is examined to check for the presence of valence shifters and the words' polarized context cluster \((c_{i,j})\) is drawn. The cluster can be represented as equation 1:

\[
(c_{i,j} = \{pwi,j,k - nb, \ldots, pwi,j,k, \ldots, pwi,j,k - na\})
\]  (1)

In equation 1, \(nb\) and \(na\) variables correspond to the n.before and n.after that the user assigned. The words in this polarized context cluster are marked neutral \((w_{i,j,k}^0\), negator \((w_{i,j,k}^n\), amplifier \((w_{i,j,k}^a\), or de-amplifier \((w_{i,j,k}^d)\). Neutral words are not given value in the equation. When polarized word is found, it is weighted \((w)\) by getting its corresponding weight in the \(polarity_dt\) argument. Then, the function further weights it based on the count of valence shifters that directly surround the polarized word \((pw)\). Punctuation locations including colons, semicolons and commas are included in the calculation of the polarized context cluster’s lower and upper limit. The reason for this is pause punctuations specify a shift in thought and words that appear before these marks are not necessarily related with words that appear after these punctuations.

The polarized word is the main value in the cluster and is altered depending on the presence of valence shifters. When amplifiers are found, the polarity is increased by 1.8. Amplifiers are treated de-amplifiers when an odd number of negators is found in the context cluster. When a de-amplifier is found, it decreases the polarity. When negations are found with amplifiers or de-amplifiers, the sign of the polarized word is reversed. To calculate negation, -1 is raised to the power of the number of negators \((w_{i,j,k}^n)\) and a value of 2 is added. The basis of this is the notion that the presence of even numbered negatives is equivalent to a positive and odd number negatives is equivalent to a negative.

Adversative conjunctions such as ‘nevertheless’, ‘though’ and ‘but’ likewise changes the context cluster. When found before the polarized word, it up-weights the cluster by equation 2:

\[
c_{i,j} up-weight = 1 + z_2 * \{|w_{adversative\ conjunction}|, \ldots, w_{i,j,k}^p\}
\]  (2)

where the default weight \((z_2)\) is 0.85 and \(|w_{adversative\ conjunction}|\) are the count of adversative conjunctions that appear before the polarized word. When found after a polarized word, adversative conjunction down-weights the cluster by equation 3:

\[
c_{i,j} down-weight = 1 + \{w_{i,j,k}^p, \ldots, |w_{adversative\ conjunction}|* - 1\}*z_2
\]  (3)

The basis of this is the idea that when an adversative conjunction is found, greater values is placed on the succeeding phrases while the value given on previous phrases are lowered.

Amplifiers and de-amplifiers were set to 0.8. Lower limit of weight for de-amplifiers is -1. The sum \((c_{i,j}^p)\) of these weighted context clusters \((c_{i,j})\) is computed. Next, the square root of the word count is calculated. Finally, the sum \((c_{i,j}^p)\) is divided by the square root as shown in equation 4:

\[
\delta_{i,j} = (c_{i,j}^p) / (\sqrt{\sum_{i,j} w_{i,j}^p})
\]  (4)

This yields an unbounded polarity score \((\delta_{i,j})\) for each sentence. To get the mean of all sentences \((\delta_{i,j})\) within a paragraph \((\delta_{i,j})\), simply take the average sentiment score in equation 5:

\[
p_i, \delta_{i,j} = 1/n * \sum_{i,j} \delta_{i,j}
\]  (5)

or utilize the default \(average_weighted_mixed_sentiment\) upweighting the negative values in a vector while also downweighting the zeros in a vector or \(average_downweighted_zero\) which simply downweights the zero polarity scores [11]. Summary statistics of the calculated sentiment scores were also observed.
One more function that was used is `highlight()`, combined with `sentiment_by()`. Combining these two functions generated an output in html highlighting sentences with green and red color. A green color denotes positive polarity while a red color denotes negative polarity. Sentences which were not highlighted are analyzed as neutral by sentimentr. We used these functions to easily observe misclassified sentences in the dataset.

3.4. Determining Correlation of Mean Sentiment Score and Actual Mean Rating
The actual mean ratings and calculated mean sentiment scores were consolidated in a csv file. We used `ggscatter()` function under `ggpubr` library in R to compute the Pearson's correlation coefficient and the significance level (or p-value) of the correlation. Pearson's correlation coefficient evaluates the statistical relationship of two continuous variables [12]. The function also generated a scatter plot illustrating the correlation of the two variables.

3.5. Validating the Sentiment Scores
Cross-tabulation of actual and predicted classes were generated using `confusionMatrix()` function. The overall accuracy and unweighted Kappa measurement were determined. Kappa measurement describes the true and predicted values' level of agreement [13]. Utilizing `mcnemar.test`, a p-value is calculated. Calculation of the overall accuracy rate was done using confidence interval of 95% for this rate and a one-sided test to check whether the accuracy is superior compared to the "no information rate", the percentage of the largest class in the dataset.

4. Results and Discussions
This section presents and discusses the findings in our experiments. Average sentiment ratings were automatically computed for each faculty. Utilizing the data of 85 teachers, the Pearson's correlation coefficient that came out from the computation is 0.73.

To determine the degree of correlation, when the coefficient value is close ± 1, at that point the correlation is considered perfect. This indicates that as one variable increments, the other variable will in general additionally increment (if positive) or decrease (if negative). A coefficient value within ± 0.50 and ± 1 means the correlation is strong. A coefficient value within ± 0.30 and ± 0.49 means the correlation is medium. A coefficient value underneath ± 0.29 means the correlation is small. If the value equals zero, then, there is no correlation [12].

This signifies that obtaining 0.73 of correlation coefficient, which is within ± 0.50 and ± 1, there is strong correlation between actual mean rating and average sentiment rating. These findings support the

![Figure 1. Scatterplot graph of actual mean ratings and average sentiment ratings.](image)
results found by Rajput, Haider, and Ghani [1] that there is positive correlation between actual mean rating and sentiment rating. The linear relationship between the two variables is illustrated in Figure 1.

In this case, there are three classes used to classify the sentences namely positive (1), negative (-1) and neutral (0). A total of 1669 sentences were analyzed in this study. Based on the analysis of the confusion matrix, 900 sentences were correctly classified as positive sentiments out of 947 True Positive sentences. Meanwhile, 518 sentences were correctly classified as negative comments out of 677 True Negative comments. Concurrently, among 45 True Neutral sentiments only 20 sentences were correctly classified.

The approach presented in this paper obtained 0.86 accuracy indicating that it is approximately 86% accurate using a confidence interval of 95% ranging from 0.8441 to 0.8778. The no-information rate of 0.5674 means approximately 57% accuracy can be achieved by always choosing the majority class label. The accuracy obtained is far higher than the no information rate indicating that our classifier is significantly better in terms of performance compared to the no-information rate. This is also showed by the p-value. The McNemar’s test P-value describes if there is significant difference between true and predicted values [14].

| Table 2. Overall Statistics. |
|-----------------------------|
| **Accuracy**                 | 0.8616 |
| **95% CI**                   | (0.8441, 0.8778) |
| **No Information Rate:**     | 0.5674 |
| **P-Value [Acc > NIR]**      | < 2.2e-16 |
| **Kappa**                    | 0.7313 |
| **Mcnemar's Test P-Value**   | < 2.2e-16 |
| **Statistics by Class:**     |         |
| **Class: -1**                | 0.7651 |
| **Class: 0**                 | 0.44444 |
| **Class: 1**                 | 0.9504 |
| **Sensitivity**              |         |
| **Specificity**              | 0.9708 |
| **Pos Pred Value**           | 0.9470 |
| **Neg Pred Value**           | 0.8583 |
| **Prevalence**               | 0.4056 |
| **Detection Rate**           | 0.3104 |
| **Detection Prevalence**     | 0.3277 |
| **Balanced Accuracy**        | 0.8680 |

A Kappa score within 0 and 0.20 is considered slight agreement. If within 0.21 and 0.40, it is fair agreement. A value within 0.41 and 0.60 is considered moderate. If within 0.61 and 0.80 it is substantial. A value within 0.81 and 1.0 is perfect agreement [13]. Since the kappa score of 0.73 is within 0.61 and 0.80, this indicates that there is a substantial level of agreement between the two values.

Predicting positive sentiments is necessary to see best practices of teachers. More so, predicting negative sentiments is crucial in this case since this require intervention measures to address the concern of students. Thus, higher value of sensitivity in class -1 (pertaining to negative sentences) is desired. Sensitivity in this context is the actual negative sentences that got predicted correctly. The overall statistics shown in Table 2 indicates that there is lower sensitivity of 0.77 for class -1 (negative sentences) as compared to 0.95 of class 1 (positive sentences). Similarly, higher positive predictive value (PPV) in class -1 is desired. Positive predictive value in this case is the confidence in the classifier that when it predicts a positive result for a class, it has the capability to distinguish successfully which belong to the class and which do not.

Specificity of class -1 (Negative) in this context is the actual not negative sentences that got predicted correctly. Table 2 indicates a higher specificity of 0.97 in class -1 (Negative) as compared to 0.82 in
class 1 (Positive). This means more not negative sentences were predicted correctly compared to not positive sentences that were predicted correctly.

Furthermore, Table 3 shows the breakdown of samples for English, Filipino and mixed sentences. Classification accuracy for English sentences got 87% and mixed English-Filipino followed at 85%. Though there is uneven distribution of samples, it can be implied that the approach has greater difficulty in analyzing Filipino sentences.

It was observed that most of the incorrectly classified English statements has words that belong to more than one parts of speech. For instance, the word ‘needs’ can be a noun or verb. Depending on how it is used, this alter the polarity of a sentence that has the word. Thus, in future research, integration of parts of speech tagging can be explored to see if it can remedy this problem. More terms, which are often used in academic environment, can also be added to the sentiment lexicon to improve the accuracy.

| Language | Total Number of Sentences | Correctly Classified |
|----------|--------------------------|----------------------|
| English  | 1340                     | 1167                 |
| Filipino | 282                      | 231                  |
| Mixed    | 47                       | 40                   |

5. Conclusion
In this paper, we presented a bilingual lexicon-based strategy for analyzing students' textual responses utilizing sentence level sentiment analysis in R. Automatic computation of sentiment polarity scores was done utilizing sentimentr. Results show that the Pearson coefficient is 0.73 which implies that there is strong correlation between real mean rating and average sentiment rating. From these discoveries, it very well may be said that as teaching performance mean rating increments, the average sentiment rating likewise increments. This demonstrates analysis of students’ comments utilizing the lexicon-based sentiment analysis technique can possibly be applied to evaluate teaching performance. Our approach got 86% accuracy. From these, it very well may be said that the methodology is encouraging as it is time-efficient and capable of analyzing the students' textual responses as compared to manual listing. With this methodology, textual responses of students can be expressed to quantitative information.

In future work, the use of POS tagging can be explored to improve sentiment analysis accuracy. Employing machine learning methods may also be considered to discover techniques and alternative approaches to sentiment classification.

References
[1] Rajput Q, Haider S and Ghani S 2016 Lexicon-based sentiment analysis of teachers’ evaluation Applied Computational Intelligence and Soft Computing Volume 2016 (London: Hindawi Publishing Corporation)
[2] Wook M, Razali N, Ramli S, Wahab N, Hasbullah N, Zainudin N and Talib M 2019 Opinion mining technique for developing student feedback analysis system using lexicon-based approach Educ. Inf. Technol. 25 pp 2549–60
[3] Rintyarna B. S, Sarno R and Faticah, C 2019 Evaluating the performance of sentence level features and domain sensitive features of product reviews on supervised sentiment analysis tasks J. Big Data 6 p 84
[4] Mukhtar N and Khan M. A 2019 Effective lexicon-based approach for Urdu sentiment analysis Artif. Intell. Rev. 53 pp 2521–48
[5] Aung K Z 2016 A lexicon based sentiment analyzer framework for student-teacher textual comments Int. J. of Scientific and Research Publications vol 6 issue 2 February 2016 ISSN 2250-3153 pp 277–80
[6] Balahadida F, Juanatas I and Fernando C 2016 Teacher's performance evaluation tool using opinion mining with sentiment analysis 2016 IEEE Region 10 Symp. (TENSYMP) Bali pp 95–8 DOI:
[7] Taj S, Shaikh B and Meghji A 2019 Sentiment analysis of news articles: a lexicon based approach 2019 Int. Conf. on Computing, Mathematics and Eng. Technologies – iCoMET pp 1-5

[8] Kundi F, Khan A, Ahmad S and Asghar M 2014 Lexicon-based sentiment analysis in the social web J. Basic. Appl. Sci. Res. 4(6) pp 238–48

[9] Reddy N, Kumar B.S, Korra B, Mohapatra S and Kumar R 2017. Sentiment analysis using Telugu sentiWordNet Int. Conf. on Wireless Communications Signal Processing and Networking (WiSPNET) - SSN College of Engineering, Chennai, India - 22-24 March 2017

[10] Yurtalan G, Koyuncu M and Turhan C 2019 A polarity calculation approach for lexicon-based Turkish sentiment analysis Turk. J. Elec. Eng. and Comp. Sci. 27 pp 1325–39 DOI: 10.3906/elk-1803-92

[11] github.com 2020 Trinker/sentimentr Retrieved on 14 March 2020 from https://github.com/trinker/sentimentr

[12] StatisticsSolutions 2020 Pearson’s Correlation Coefficient Retrieved on 13 March 2020 from https://www.statisticssolutions.com/pearsons-correlation-coefficient/

[13] Chen Y S 2019 Interpretation of Kappa Values Retrieved from https://towardsdatascience.com/interpretation-of-kappa-values-2acd1ca7b18f

[14] Ashcroft M 2016 The Essentials of Data Analytics and Machine Learning Retrieved from https://courses.edsa-project.eu/pluginfile.php/1330/mod_resource/content/0/Module%2010%20-%20Analysis%20of%20Classification%20Models_V1.pdf