Section 1. Higher Education

Dr., Tzacheva Angelina,
teaching associate professor
University of North Carolina at Charlotte USA
E-mail: aatzache@uncc.edu

Jaishree Ranganathan,
Ph D., student
University of North Carolina at Charlotte USA
E-mail: jrangan1@uncc.edu

EMOTION MINING FROM STUDENT COMMENTS A LEXICON BASED APPROACH FOR PEDAGOGICAL INNOVATION ASSESSMENT

Abstract: Course evaluation provided by student's play a major role in a wide range of factors that include suggestions on areas of improvement in terms of teaching, available resources, study environment, and student assessment techniques. These evaluations are collected in both quantitative and qualitative forms. The quantitative feedbacks include a Likert-type scale in which responses are scored along a range, to capture the level of agreement and disagreement. Whereas the qualitative feedbacks provide an open portal for the students to convey their feelings, thoughts or opinion about the course, instructor and assessments in a more general way. The qualitative data is in the form of textual comments which can be processed to mine student's emotional feeling and gain more intellectual insights. In this work we focus on qualitative student feedbacks through text mining and sentiment analysis. We analyze the efficiency of Active Learning methods Light Weight teams and Flipped Classroom. Results show the implementation of these methods is linked with increased positivity in student emotions.

Keywords: Data Mining; Education; Emotion Mining; Flipped Classroom; Light Weight Teams; Visualization.

Introduction:

Student evaluation of teaching is an important element in the process of evaluating and improving instruction in higher education as described by Zabaleta [1]. These evaluations help not only in teaching improvements but also in some of the decisions like future employment, retention, and promotion of faculty. It is now-a-days common in almost any educational institution to collect end of course evaluation, which allows students to express their feelings or opinion about the instructor. These evaluations are collected at the end of course typically end of semester. There are basically two types of question format in the evaluation system: Quantitative and Qualitative. Quantitative questions are Likert-type items which the students can respond in the scale of 1 to 5, starting

with Strongly Agree – 1, Agree – 2, Neutral – 3, Disagree – 4, and Strongly Disagree – 5. Qualitative questions are open ended questions where students can write their opinion, and/or thoughts in a free style manner. According to author Clayson [2] since the 1970's the application of student evaluation in teaching has become nearly universal.

Data Mining is one of promising fields which involves the practice of searching through large amounts of computerized data to find useful patterns [3]. These patterns are then utilized by analysts to find interesting measures and apply strategies to improve the current methodology or practices. According to author's Spooren et al. [4] there are three main purpose for which student evaluations are used as follows: a) improve teaching methodology and/or quality; b) serve as input for tenure/ promotion decisions, and; c) Demonstrate the evidence of institutional accountability in terms of resources and environment provided. Mining this kind of educational data is one of the important areas of research which is gaining importance in recent years due to increase in the demand of quality education and the demography of students attending higher education. Most of the students in recent years are Millennials and their mindset towards education is different which requires better understanding from University and the Instructors in order to provide a better experience in education.

In recent years there is an increase in the need for understanding what is said about a element. For instance, in an online store, customer reviews about a product – where customers convey their opinion about the quality and usefulness of the product and how well it suits their expectation. These kind of reviews helps business analyst improve their marketing strategies and apply to the quality of the products. Understanding people's feeling or emotion is a separate area of research which is called Sentiment Analysis.

The word Emotion dates to 1570's, derived from old French 'emouvoir' meaning 'stir up' according to online Etymology dictionary. Scientific research in understanding Human Emotion's dates to 1960's.

For instance, Ekman [5] studied human emotions and their relation to facial expressions. According to Ekman there are six basic emotions 'anger', 'disgust', 'fear', 'joy', 'sadness', and 'surprise'. Similarly, there are other scientists who proposed emotion theories, Author James [6] and Plutchik [7]. In [8], the authors discuss different basic emotion models proposed by theorists since 1960. In this work we use the National Research Council – NRC Emotion lexicon [9; 10].

In this paper we focus on mining student feedback collected from the end-of-semester course evaluations, in particular the qualitative results and identify student's emotion to understand whether incorporation of Light Weight teams [11; 12], and Flipped Classroom techniques [13] helped students during the course for the time period 2013 to 2017.

Reminder of this paper is organized as follows: section II talks about the related work in the area, section III the methodology in data extraction and emotion labeling followed by Experiments and results in section IV and Discussion and Conclusion in section V and VI respectively.

Related Work:

In this section we review studies that have been done in the area of analyzing student evaluations, including text and quantitative data.

Authors Kim et al. [14] perform Sentiment Analysis on the ratings and textual responses of student evaluation of teaching. They automatically rate the textual response as one of the three categories 'positive', 'negative', and 'neutral'. In which they have compared the performance of categorical model and dimensional model where 'joy' and 'surprise' are positive class, 'anger', 'fear' and 'sadness' are negative class respectively. In their work they have utilized two emotion lexicons WordNet-Affect and ANEW for the sentiment classification tasks. The following five approaches are modeled for automatic classification of three sentiments 'positive', 'negative', and 'neutral': a) Majority Class Baseline (MCB); b) Keyword Spotting (KWS), c) CLSA – LSA based categorical classification; d) CNMF – NMF based categorical classifica-

tion, and; e) DIM – Dimension based estimation. It is shown in terms of precision, recall and f-measure that NMF based categorical and dimensional models have a better performance than other models.

Typically, in an end-of-course evaluation the students do not benefit to see the actions taken as they move on from the section after that semester. In order to overcome it is required to obtain prompt feedback from students to instructors and necessary actions can be taken during the course. Authors Leong et al. [15] propose the use of short message service (SMS) for student evaluation and explore the application of text mining in particular Sentiment Analysis ('positive' and 'negative') on SMS texts. They show the positive and negative aspects of lecture in terms of the conceptual words extracted and text link analysis visualization.

Similar to [15] authors Altrabsheh et al. [16] explore approaches for real time feedbacks. This work discusses how feedback is collected via social media such as Twitter and apply Sentiment Analysis to improve teaching called as *Sentiment Analysis for Education (SA-E)*. This system collects data from Twitter where the students provide their feedback. The text data after pre-processing and extracting features including: term presence and frequency, N-gram position, part-of-speech, syntax, and negation. Later the text is analysed via Naive Bayes and/or Support Vector Machine which categorizes the whole post as either `positive' or 'negative'.

Authors Jagtap et al. [17] perform Sentiment Analysis on student feedback data classifying into 'positive' and 'negative' categories. They combine Hidden Markov Model (HMM) and Support Vector Machine (SVM) and use a hybrid approach for sentiment classification. Though they have concluded that applying advance feature selection method combined with hybrid approach work well for complex data, their works did not show the results of classification model for validation.

Authors Rajput et al. [18] apply text analytics methods on student's feedback data and obtain insights about teacher's performance with the help of tag clouds, and sentiment score. In this work the authors use sentiment dictionary Multi-Perspective Question Answering (MPQA) [24] to find words with positive and negative polarity. By combing the word frequency and word attitude the overall sentiment score for each feedback is calculated. Finally, they have compared the sentiment score with Likert scale-based teacher evaluation and conclude that Sentiment score with word cloud provide better insights than Likert-scale results.

In this paper we propose analyzing the qualitative end-of-course teacher evaluations with fine grained emotions such as 'anger', 'trust', 'sadness', 'joy', 'anticipation', 'fear', and 'disgust' with the help of National Research Council – NRC Emotion lexicon and combing the word frequency and sentiment score to determine the overall sentiment – emotion associated with student comments.

Methodology:

This section details the approach used in this paper to process the student evaluation data. The following are the steps involved in the experimental framework. Data collection, data extraction, pre-processing, Emotion labeling, visualization. The overall methodology is shown in (Figure 1).
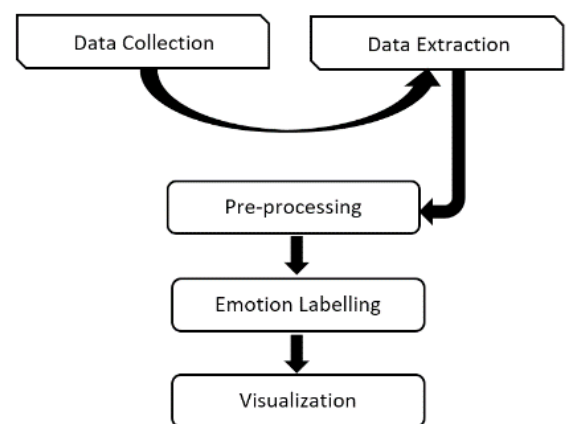

Figure 1. Methodology

Data Collection:

The data for this study is collected from the Web-Based course evaluation system by UNC Charlotte. This system is administered by a third-party

Campus Labs. In assistance with UNC Charlotte Center for Teaching and Learning, Campus Labs collect the student feedback for course evaluations. The student feedbacks for an instructor is collected for the terms of 2013 to 2017 including Fall, Spring and Summer sections of various courses handled by the instructor. We collect the html files from Campus Labs website for each of the semester. Next, we process the data as described in the Data Extraction subsection below. This data includes both quantitative and qualitative results. For this study we used qualitative feedback mainly focusing on Sentiment Analysis. Sample qualitative data shown in (Table 1). The (Table 2). shows the list of semesters for which the data is collected.

Table 1. – Sample Student Feedback – Qualitative

S. NO	Top
1.	Easily available to communicate with if needed
2.	The course has a lot of valuable information
3.	Get rid of the group project
4.	There was no enthusiasm in the class. The instructor should make the class more lively and interactive.
5.	Best professor

Table 2. – List of Semesters – Student Feedback

Year	Semester
2013	Spring, Summer, Fall
2014	Spring, First Summer, Second Summer, Fall
2015	Spring, First Summer, Second Summer, Fall
2016	Spring, Spring Midterm, First Summer, Second Summer, Fall
2017	Spring, First Summer, Second Summer, Fall

Table 3. – Sample Data Extracted

Year	Term	Course	Question	Comments
2014	Fall 2014	Operating Systems and Networking	Please list outstanding strengths of the course and/or instructor	Easily available to communicate with if needed
2014	Fall 2014	Operating Systems and Networking	Please list outstanding strengths of the course and/or instructor	The course has a lot of valuable information
2014	Fall 2014	Operating Systems and Networking	Please provide other observations, comments, or suggestions	Get rid of the group project
2017	Fall 2017	Cloud Comp for Data Analysis	Please suggest areas for improvement of the course and/or instruction method	There was no enthusiasm in the class. The instructor should make the class more lively and interactive
2017	Fall 2017	Cloud Comp for Data Analysis	Please list outstanding strengths of the course and/or instructor	Best professor

Data Extraction:

After the data collection from Campus Labs, jsoup [19] a Java library is used to process the html files and extract the comments. The following fields are extracted from the html file: Year, Term, Course, Questions, Comments. Sample data is shown in Table 3. The data extracted consists 959 records with the five attributes as mentioned in (Table 3).

Pre-Processing:

Pre-processing is one of the important steps in handling text data. This involves removal of noisy and unwanted parts from the text. In this work the Python Natural Language Toolkit (NLTK) [20] is used to work with student evaluation data. The following steps are involved in pre-processing of student course evaluation comments: Tokenization, lower case, stop words removal.

Tokenization:

Tokenization is the process of splitting the text or sentence into words. In specific it is the task of chopping character sequences into pieces called tokens (words) and removing certain characters like punctuation. An example is shown in (Figure 2).

Input: "She is good at reading powerpoints, I guess."

Output: She is good at reading powerpoints I guess

Figure 2. Tokenization

Lower Case:

Natural language text written by human beings contains both lower case and upper case. In terms of processing this kind of text using a machine requires all the text to be in same case for better performance. This step changes the text to lower case.

Stop Words Removal:

Some of the words in English language are frequently used in order to make the sentence more complete in terms of grammar. These words are generally not much useful in terms of the context of the sentence in most of the cases. For instances words like 'am', 'is', 'was', 'are' etc. There is list of stop words available in the Python Natural Language Toolkit (NLTK) [20] corpus which is used as part of this stop words removal step.

In the pre-processing step, certain comments which are not valid are removed for instance comments with only 'n/a', 'NA', etc. The pre-processed dataset contains close to 700 records in the dataset.

Emotion Labeling:

After Data extraction and pre-processing the next important step is labeling the data – student feedback comments with different types of Emotion. We use the National Research Council – NRC Lexicon [9], [10] for this purpose. NRC Emotion lexicon is a list of English words and their associations with eight basic emotions (anger, fear, anticipation, disgust, surprise, trust, joy, and sadness) and two sentiments (positive and negative). The Annotations in the lexicon are at WORD-SENSE level. Each line has the format: <Term> <**AffectCategory**> <**AssociationFlag**> as shown in (Figure 3). A tree map of the lexicon with words containing Flag as 1 for each of the respective emotion filtered is shown in (Figure 4).

Each of the student comments is processed and if a match to word is found then the score is incremented accordingly based on the Flag value in lexicon, here if a word is present twice then automatically based on the frequency score for that particular emotion will be incremented. After the entire comment is processed the Emotion which has the highest score is assigned as the final Emotion with respect to that student comment. As part of Emotion labeling if the final emotion score is zero then those records are omitted from the dataset.

Visualization:

This paper mainly focuses on identifying if the students are feeling better in a way the course is delivered with changes including Light Weight teams,

flipped class room, and active learning methodologies. After labeling the students feedback with appropriate Emotion class, the data is used to visualize the results over the years 2013 to 2017 and results are analyzed. For visualization Tableau software [21] is used. Visualization is a powerful tool for exploring large data, both by itself and coupled with data mining algorithms [22].

<term><tab><AffectCategory><tab><AssociationFlag>

<term>: is a word for which emotion associations are provided;

<AffectCategory>: is one of eight emotions (anger, fear, anticipation, trust, surprise, sadness, joy, or disgust) or one of two polarities (negative or positive);

<AssociationFlag>: has one of two possible values: 0 or 1. 0 indicates that the target word has no association with affect category, whereas 1 indicates an association.

Figure 3. NRC Emotion Lexicon – Word Level Annotation
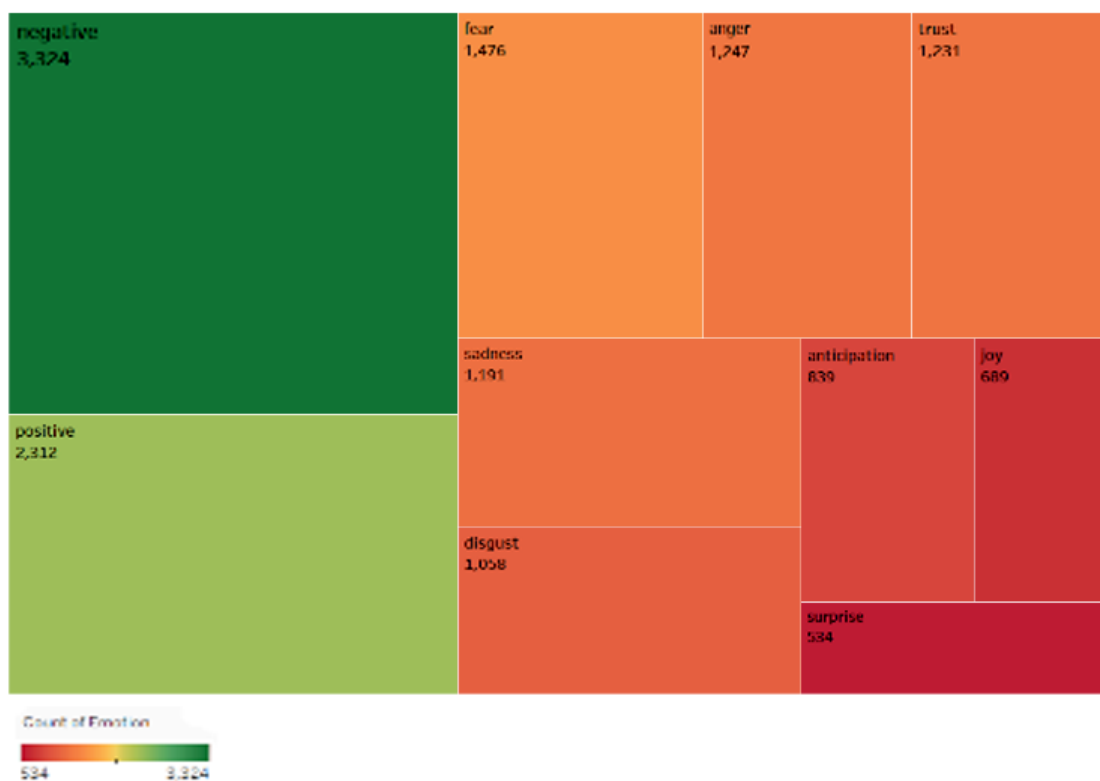
Figure 4. Tree Map – NRC Emotion Lexicon

Experiments and Results:

In this section we describe our experiments and results. The data for this study is collected from UNC Charlotte Campus Labs website. Sample qualitative data shown in (Table 1). The data extracted consists 959 records with the five attributes as mentioned in (Table 3). The pre-processed dataset contains close to 700 records in the dataset. For labeling the data – student feedback comments with different types of Emotion. We use the National Research Council – NRC Lexicon [9; 10].

We conduct two sets of experiments one which includes the positive and negative polarity along with basic emotions 'anger', 'trust', 'fear', 'sadness', 'disgust', 'anticipation', 'surprise', and 'joy', other only with the basic emotions. The experiments are separated in such way because the lexicon used contains most of the words tagged as positive and negative as shown in (Figure 4).

Experiment 1 – Labeled with Basic Emotion and Polarity:

In Experiment 1, the pre-processed data is passed to the system which finds the word associated with 8 basic emotions and the polarity for each of the stu-

dent feedback response. After which the scores are calculated based on the frequency of each emotion and polarity related words. The sentiment that has highest score is assigned as overall emotion/polarity. The results are shown on a temporal basis from 2013 until 2017 on the X-axis and the count of each emotion on the Y-axis in (Figure 5). It is observed that emotion 'trust' and polarity 'positive' has a growing trend through the time. Similarly, we see that `anticipation' was high during the year 2014 which gradually decreased in the year 2017. These changes are attributed towards active learning methodology implemented in the year 2016 and 2017.
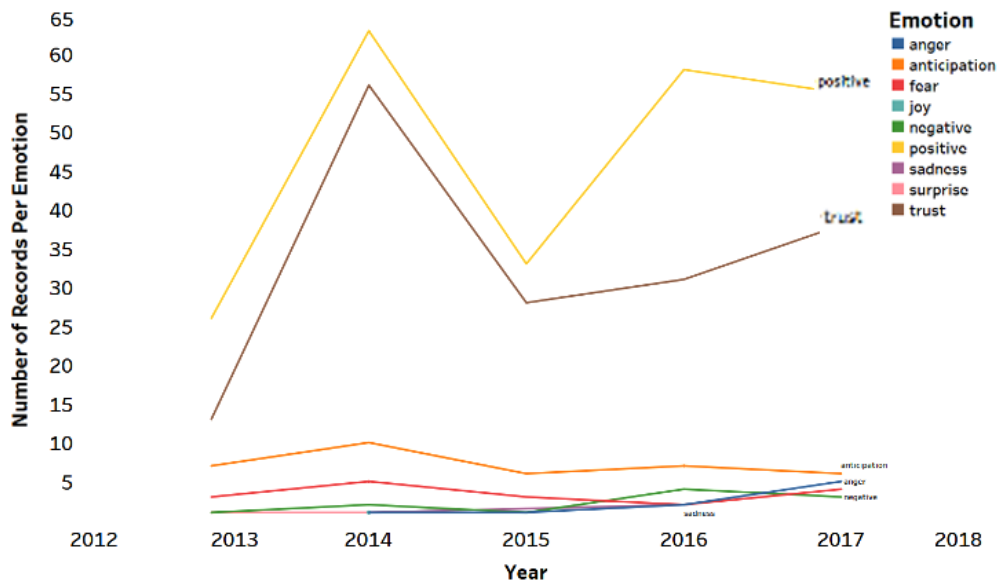

Figure 5. Experiment 1 – Basic Emotion and Polarity.

Experiment 2 – Labeled with Basic Emotions:

In Experiment 2, the pre-processed data is passed to the system which finds the word associated with 8 basic emotions for each of the student feedback response. After which the scores are calculated based on the frequency of each emotion related words. The sentiment that has highest score is assigned as overall emotion. The results are shown on a temporal basis from 2013 until 2017 on the X-axis and the count of each emotion on the Y-axis in Figure 6. The results for this experiment is almost the same as Experiment 1, without the two polarities 'positive' and 'negative'. It is observed that emotion 'trust' has a growing trend

through the time. Similarly, we see that 'anticipation' was high during the year 2014 which gradually decreased in the year 2017. In Experiment we observe emotion 'joy' for the year 2016 when actually active learning methodology was started in the classes. But the count of the emotion 'joy' is low compared to others in the data.

Sentiment Analysis and Emotion Detection in Student Evaluations – Word Cloud:

Word Cloud is a text summarization, which shows the most frequently occurring words in a text, with the largest font. Word Cloud is helpful to learn about the number and kind of topics present in the text [23].
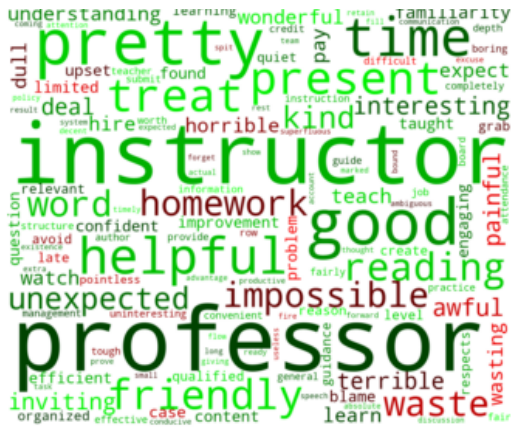
Figure 6. Experiment 2 – Basic Emotion

In this work we use the Word Cloud package in Python to create Word Clouds using the emotional words from the student evaluation data. During the emotion labeling step for each of the student feedback, the emotional words are recorded separately for each of the eight emotion and the positive and negative polarities. To form word-cloud the list of words from the following emotions 'anger', 'fear', 'sadness', 'disgust', and 'negative' are taken as negative word list from the NRC Emotion Lexicon [9;10]. These words appear in 'red' color in the word cloud. The positive words are words that denote the following emotion 'joy', 'trust', 'anticipation', and 'positive' polarity appear in grey scale. The most frequently occurring positive Words are shown in green color.

We observe that the year 2014 and 2015 have more negative words including 'problem', 'waste', 'disappointed', 'awful', 'painful' and others as shown in (Figure 8) and (Figure 9). In 2017, the (Figure 11). shows more frequency of positive words like 'helpful', 'resources', 'good', 'information'. In 2017 Active Learning methods were implemented in the courses, including Light Weight Teams [11; 12], and Flipped Classroom [13]. We show that occurrences of negative emotion words in the Figure 11. like 'terrible' have decreased since 2017.

Figure 7. WordCloud-2013. Most frequent word appears with largest font. Negative words in red. Positive words in green

Figure 8. WordCloud-2014. Most frequent word appears with largest font. Negative words in red. Positive words in green

Figure 9. WordCloud-2015. Most frequent word appears with largest font. Negative words in red. Positive words in green

Figure 10. WordCloud-2016. Most frequent word appears with largest font. Negative words in red. Positive words in green

Figure 11. WordCloud-2017. Most frequent word appears with largest font. Negative words in red. Positive words in green

Therefore, we claim that the implementation of Light Weight Teams and Flipped Classroom Active

Learning methods increase positive emotions among students and improve their learning experience.

Discussion:

In this work we use the NRC emotion lexicon [9; 10] and label each student feedback with appropriate emotion based on the overall score of emotional word frequency. We see that the words like 'examination', 'presentation', 'subject' are normal terms that students use to describe a course. These words in general are considered negative but not in educational domain as they are typical to explain any course requirements. It is identified that general purpose lexicon does not suit the educational domain directly but require some changes.

Also, we see some false positives in the emotion labeling for the contents as follows: For instance, the following comment *'Don't talk at us for 2 and a half hours. The class would do well to integrate clicker questions and discussion'* is assigned a positive emotion according to the methodology adopted. This is because of the presence of words like 'well', 'talk', 'discussion'. Another example *'Please change the test structure to actually test the student's knowledge and assign more programming projects.'* is assigned a positive emotion due to the presence of words like 'structure', 'knowledge'.

Conclusion:

In this work we perform sentiment analysis, and emotion detection on the qualitative feedback provided by students in course evaluations. We identify eight basic human emotions: 'anger', 'fear', 'joy', 'surprise', 'anticipation', 'disgust', 'sadness', and 'trust' along with the two sentiment polarities 'positive' and 'negative'. We use these emotions to analyze and assess the impact and effectiveness of Active Learning methods incorporated in the classroom during the years 2016 and 2017, compared to previous years. Active Learning methods were initiated in 2016, and implemented in 2017, in the courses including Light Weight Teams [11; 12], and Flipped Classroom [13]. Results show evidence that words associated with positive emotions, and trust have increased in the recent years

compared to 2014. At the same time, occurrences of negative emotion words in the (Figure 11). have decreased. Therefore, we claim that the implementation of Light Weight Teams and Flipped Classroom Active Learning methods increase positive emotions among students and improve their learning experience. In the future we plan to extend this work, by analyzing more Active Learning pedagogy methods such as gamification. We also plan to focus on women and minorities in computing discipline.

Acknowledgements

References:

1. Zabaleta F. The use and misuse of student evaluations of teaching. Teaching in Higher Education, – Vol. 12. – No.1. 2007. – P. 55–76.
2. Clayson D. E. Student evaluations of teaching: Are they related to what students learn? A meta-analysis and review of the literature. Journal of Marketing Education, – Vol. 31. – No. 1. 2009. – P. 16–30.
3. Dictionary M. W. (2002). Merriam-Webster. On-line at URL: http://www. mw. com/home. htm
4. Spooren P., Brockx B., & Mortelmans D. On the validity of student evaluation of teaching: The state of the art. Review of Educational Research, – Vol. 83. – No. 4. 2013. – P. 598–642.
5. Ekman P. An argument for basic emotions. Cognition & emotion, – Vol. 6. – No. 3–4. 1992. – P. 169–200.
6. James W. What is an emotion? Mind, – Vol. 9. – No. 34. 1884. – P. 188–205.
7. Plutchik R. Emotions and psychotherapy: A psychoevolutionary perspective. In Emotion, psychopathology, and psychotherapy, – P. 3–41.
8. Ortony A., & Turner T. J. (1990). What's basic about basic emotions? Psychological review, – Vol. 97. – No. 3. 1990. – P. 315.
9. Mohammad S. M., & Turney P. D. Crowdsourcing a word–emotion association lexicon. Computational Intelligence, – Vol. 29. – No. 3. 2013. – P. 436–465.
10. Mohammad S. M., & Turney P. D. Emotions evoked by common words and phrases: Using Mechanical Turk to create an emotion lexicon. In Proceedings of the NAACL HLT 2010 workshop on computational approaches to analysis and generation of emotion in text, Association for Computational Linguistics, 2010. Junepp. – P. 26–34.
11. Latulipe C., Long N. B., & Seminario C. E. Structuring flipped classes with lightweight teams and gamification. In Proceedings of the 46th ACM Technical Symposium on Computer Science Education, 2015. – P. 392–397.
12. MacNeil S., Latulipe C., Long B., & Yadav A. Exploring lightweight teams in a distributed learning environment. In Proceedings of the 47th ACM Technical Symposium on Computing Science Education, 2016. – P. 193–198.
13. Maher M. L., Latulipe C., Lipford H., & Rorrer A. Flipped classroom strategies for CS education. In Proceedings of the 46th ACM Technical Symposium on Computer Science Education, 2015. – P. 218–223.
14. Mac Kim S., & Calvo R. A. Sentiment analysis in student experiences of learning. In Educational Data Mining. 2010.
15. Leong C. K., Lee Y. H., & Mak W. K. Mining sentiments in SMS texts for teaching evaluation. Expert Systems with Applications, – Vol. 39. – No. 3. 2012. – P. 2584–2589.
16. Altrabsheh N., Gaber M., & Cocea M. SA-E: sentiment analysis for education. In International Conference on Intelligent Decision Technologies, – Vol. 255. 2013. Junepp. – P. 353–362.

17. Jagtap B., & Dhotre V. SVM and HMM based hybrid approach of sentiment analysis for teacher feedback assessment. International Journal of Emerging Trends & Technology in Computer Science (IJETTCS),– Vol. 3.– No. 3. 2014.– P. 229–232.

18. Rajput Q., Haider S., & Ghani S. Lexicon-Based Sentiment Analysis of Teachers' Evaluation. Applied Computational Intelligence and Soft Computing, 2016.– 1 p.

19. Hedley J. (2009). jsoup: Java html parser. 2009. – P. 11–29. [2015–06–12] URL: http://jsoup. org.

20. Bird S., Klein E., & Loper E. Natural language processing with Python: analyzing text with the natural language toolkit. " O'Reilly Media, Inc." 2009.

21. Hanrahan P. Tableau software white paper-visual thinking for business intelligence. Tableau Software, Seattle, WA. 2003.

22. Hanrahan P., & Stolte C. U. S. Patent No. 7, 800, 613. Washington, DC: U. S. Patent and Trademark Office. 2010.

23. Heimerl F., Lohmann S., Lange S., & Ertl T. Word cloud explorer: Text analytics based on word clouds. In System Sciences (HICSS), 2014. 47th Hawaii International Conference on (P. 1833–1842). IEEE.

24. Wiebe J., Wilson T., & Cardie C. Annotating expressions of opinions and emotions in language. Language resources and evaluation,– Vol. 39.– No. 2–3. 2005.– P. 165–210.