

## Natural language processing for analysis of student online sentiment in a postgraduate program.

Truc D Pham<sup>1</sup>, Darcy Vo<sup>1</sup>, Frank Li<sup>1</sup>, Karen Baker<sup>1</sup>, Binglan Han<sup>1</sup>, Lucie Lindsay<sup>1</sup>, Mohsen Pashna<sup>2</sup>, Rich Rowley<sup>2</sup>

<sup>1</sup>The Mind Lab, Auckland, New Zealand

<sup>2</sup>Tech Futures Lab, Auckland, New Zealand

### Abstract

Higher education institutes are continually looking for new and better ways to support and understand the learning experience of their students. One possible option is to use sentiment analysis tools to investigate the attitudes and emotions of students when they are interacting on social media about their course experience. In this study, we analysed the social media posts, from a closed programme-based community, of more than 300 students in a single programme cohort by processing the dataset with the Google cloud-based Natural Language Processing API for sentiment analysis. The sentiment scores and magnitudes were then visualised to help explore the research question ‘**How does a natural language processing tool help analyse student online sentiment in a postgraduate program?**’ The results have provided a better understanding of students’ online sentiment relating to the activities and assessments of the programme as well as the variation of that sentiment over the timeline of the programme.

Key words: Sentiment analysis; online learning; online interaction; postgraduate studies

### Introduction

Educational institutions globally are increasingly interested in using different innovative approaches to gauge student attitudes towards their learning experience beyond the traditional student feedback forms. Sentiment analysis of student social media posts using Natural Language Processing Tools is one option that may be used to achieve this goal. For the study reported in this article, the Google Cloud Natural Language API was chosen as a sentiment analysis tool for its ability to use both score and magnitude as the means of analysing student self-expression online. This paper recounts our experience of applying this tool to our Google Plus (G+) community to determine its applicability in a blended postgraduate course that contained both face to face components and online learning. This research investigated student sentiment in online postings, using predetermined categories, over periods of time throughout the programme. We explored how a natural language processing tool can be used to analyse student online sentiment in a postgraduate programme. In the rest of this article, we present a range of data and analytical outcomes that provide some insights into the potential of sentiment analysis on gauging student attitudes to their course experience.

### Literature review

Sentiment analysis (SA) is “the computational treatment of opinion, sentiment, and subjectivity in text” (Pang & Lee, 2008.p.5). In sentiment analysis research, people’s opinions expressed in the text towards entities such as products, services, organisations, individuals, issues, events, topics, and their attributes are identified and classified into different sentiments (Liu, 2012; Rani & Kumar, 2017). The basic sentiment analysis model consists of the collection, analysis and pre-processing of data, sentiment classification and visualisation (Ozturk, Cicek & Ergul, 2017, p. 753). Sentiment classification attempts to automatically determine whether a text represents a positive or negative emotion of the writers (Kim & Calvo, 2010; Thelwall, 2016).

There has been a huge increase in the number of studies in sentiment analysis since 2015 (Mäntylä, Graziotin & Kuutila, 2018). Sentiment analysis research initially focuses on review-related websites (Pang & Lee, 2008;

Mäntylä, Graziotin & Kuutilla, 2018 ) and public opinion towards political figures or situations (Mäntylä, Graziotin & Kuutilla, 2018). Recently, with the growth of social media use, most papers draw data from online social networking platforms such as Twitter and Facebook (Liu, 2012; Mäntylä, Graziotin & Kuutilla, 2018).

In education, researchers are interested in understanding student learning experiences and their attitudes towards a course or programme, and student feedback is often the primary source of information to do so. There are a number of studies investigating the data extracted from online student surveys (Kim & Calvo, 2010; Nitin, Swapna & Shankaraman, 2015; Balahadia, Fernando & Juanatas, 2016; Dhanalakshmi & Bino, & M. Saravanan, 2016; Rani & Kumar, 2017; Aung, K & Myo, 2017). When using sentiment analysis techniques to explore textual survey responses, student sentiments can be quickly detected (Kim & Calvo, 2010; Rani & Kumar, 2017). However, while student feedback provides valuable information about the student learning experience and attitudes, the input can only show the overall emotions at the end of the course or programme. Student emotions may change during the course of their learning. In contrast, in an online professional community, posts are put up more frequently and throughout their learning. Analysing this type of student online interaction data provides an understanding of one of the pieces of a whole picture about the student learning experience. Studies such as that of Altrabsheh, Cocea and Fallahkhair (2014), Sivakumar and Reddy (2017) and Ozturk, Cicek and Ergul ( 2017) sourced data from tweets, while Colace, De Santo and Greco (2014) extracted comments in chats and forums on Moodle to look at student sentiments towards the teaching and learning.

There are different approaches to sentiment analysis. Liu (2012) categorises sentiment analysis approaches according to the level of analysis, which is generally at three levels: document level, sentence level, entity and aspect level. Sivakumar and Reddy's (2017) study analysed online student feedback by looking at the semantic relatedness of aspect word and student opinion sentence. Ozturk, Cicek and Ergul (2017) fetched Tweets related to their university's open and distance education and extracted sentiments at the sentence level. In their research, Colace, De Santo and Greco (2014) developed a Mixed Graph of Terms as a tool for the sentiment classification of documents to detect student feelings about teaching topics. Mäntylä, Graziotin and Kuutilla (2018) group research studies of sentiment analysis according to the data-analysis methods into machine learning, natural language processing and sentiment analysis specific methods. Most studies in sentiment analysis in education used specific machine learning classification techniques such as Naïve Bayes classification algorithm (Sivakumar & Reddy, 2017; Ozturk, Cicek & Ergul, 2017) and Support Vector Machines (Altrabsheh, Cocea & Fallahkhair, 2014). When comparing the sentiment analysis of Natural Language Processing from different services including Stanford University, Google, Microsoft and IBM, NLTK, OpenAI, Apache OpenNLP, LexisNexis, Amazon in their research, Vasilescu, Suci, Pasat (2019) found that Google service provides the best results. However, there has not been any research that uses it to analyse the sentiment of student online interaction. Our study used the application to address the research question "How does a natural language processing tool help analyse student online sentiment in a postgraduate programme?".

## Background

The programme investigated in this study is a part-time postgraduate professional development programme for secondary, primary or tertiary teachers.

As busy professionals, these teachers have heavy responsibilities both at school and at home with families. The biggest challenge for a teacher to attend part-time postgraduate study is how to balance these responsibilities and study to achieve their learning goals. Compared to full-time tertiary students who can attend face-to-face lectures and meet and collaborate with classmates every day, these part-time professional learners conduct their study mainly online - including seeking help from staff and peer support via online discussion forums.

The 32-week programme consists of four courses including two face-to-face courses, which are 'Digital & Collaborative Learning in Context' and 'Leadership in Digital & Collaborative Learning', and two online courses, which are 'Applied Practice in Context' and 'Research and Community Informed Practice'. The abbreviations used in the G+ community for the four courses are DIGITAL for Digital & Collaborative Learning in Context course, LEADERSHIP for Leadership in Digital & Collaborative Learning, PRACTICE for Applied Practice in Context, RESEARCH for Research and Community Informed Practice.

To support our students’ online learning, we created a single national closed G+ Online Community for each intake. We encourage the students to share their own learning journey with peers, seek collaborative opportunities, and gain support from peers and staff, for example, they can start or contribute to a discussion topic, post queries and respond to others. We also created different categories in each G+ community to organize the posts. However, the joining and participating in online discussions on this platform is not compulsory, students can also join professional forums on other platforms such as Twitter, Facebook or Pinterest. The two online courses during the second half of the programme require students to submit evidence of online participation for assessments. The grade for the evidence of online participation contributes to only a small proportion of the overall grade for each course.

The data that we used for this study is retrieved from the G+ community of the July 2017 intake. This intake started from 17 July 2017 and ended on 25 March 2018, but students’ online engagement with this G+ community lasted far beyond the official end of the program. The categories on the G+ community for the first 16 weeks, during which there are weekly face-to-face sessions, are ‘Introduction’, ‘Technical Help’, ‘Interesting Links’, ‘Student reflections’, ‘In Class Tasks’, and ‘Assessment Related Discussions’. As students needed more support once they started their self-directed and self-paced learning for their online courses in the second 16 weeks, we created more categories under this G+ community to cater to their needs. We included ‘RESEARCH: Literature review’, ‘RESEARCH: Teacher Inquiry Project Plan’, and ‘RESEARCH: Online Participation’. Each of these categories was for an assessment in the ‘Research and Community Informed Practice’ course. For PRACTICE course, we created separate categories for each week, as each week covered a different topic and aligned with a separate activity for assessment. The names for these categories start with ‘PRACTICE week .....’, for example, ‘PRACTICE week 28: Influence of Law and Ethics’. In the second half of the programme, we also provided weekly an hour live chat with students on the G+ community to provide extra help. During these sessions, the facilitators answered students’ inquiries in real-time. The names for these categories are ‘RESEARCH: LIVE Q&A SESSION’ and ‘PRACTICE: LIVE Q&A SESSION’.

## Methodology

Google Natural Language Processing (Google NLP) has been used for sentiment analysis in this project. In order to understand the text, from a range of approaches offered by Google’s NLP API, the Sentiment Analysis option has been used. This service estimates the ‘emotional opinion’ of the text to determine the attitude of the writer with two numerical parameters which are ‘score’ and ‘magnitude’ for the whole given text and every single sentence in the text.

According to Google’s NLP, the ‘score’ is a real number in the range -1 to 1, and represents the emotional understanding of text. A text or document with positive emotions will have the score in the range 0 to 1, and text or documents with negative emotions will have a score in the range -1 to 0. When the score is close to zero, the emotion can be considered as being neutral. For example, Google’s NLP suggests that text or documents can be considered as emotionally neutral if the score is in the range from -0.25 to +0.25. Table 1 shows the range of values of scores for being negative, neutral, and positive in emotion.

**Table 1: Range of value for the emotional score**

-1.0 -0.25	-0.25 0.25	0.25 1.0
Negative	Neutral	Positive

Google’s NLP use ‘magnitude’ to indicate the summation of every emotional expression in the text, and represents the amount of emotions. The ‘Magnitude’ is the absolute value and in the range from 0 to positive infinite and usually has a direct proportional relationship with the length of the text. Magnitude helps to distinguish documents with combined positive and negative emotions in contrast with totally neutral documents. Table 2 below tabulates example score and magnitude interpretations.

**Table 2:** Examples of emotion Interpretation with different values of Score and Magnitude (Google, 2020b).

Sentiment	Sample Values
Clearly Positive*	"score": 0.8, "magnitude": 3.0
Clearly Negative*	"score": -0.6, "magnitude": 4.0
Neutral	"score": 0.1, "magnitude": 0.0
Mixed	"score": 0.0, "magnitude": 4.0

Table 3 shows a few comments created by the authors and the respective sentiment score and magnitude returned by Google Cloud Platform (Google, 2020a) at both Document-Level analysis and Sentence-Level analysis as some examples of how the cloud-based tool works. These comments have relevant context to the programme. As the Ethics application of the project allows anonymous data collection only, authentic comments from the students can not be presented here.

**Table 3:** Examples of relevant comments and their sentiment analysis by Google Cloud Platform (Google, 2020a).

No	Comment	Level of Sentiment Analysis	Sentiment Score	Sentiment Magnitude
1	Thank you for your feedback, the literature is really helpful. They also show me other relevant sources.	Entire document	0.6	1.2
		Thank you for your feedback, the literature is really helpful.	0.9	0.9
		They also show me other relevant sources.	0.3	0.3
2	I am really frustrated, I can't find any relevant literature for this topic. I think I need to change the direction of my work for this assessment.	Entire document	-0.2	1.4
		I am really frustrated, I can't find any relevant literature for this topic.	-0.9	0.9
		I think I need to change the direction of my work for this assessment.	0.4	0.4

In this paper, text with a score between 0.25 and 1 is considered as containing positive sentiment. Text with a score between -1 and -0.25 is considered as containing negative sentiment. Text with a score in the range [-0.25, +0.25] and magnitude in the range [0, 1] is considered neutral. Text with a score in the range [-0.25, +0.25] and magnitude greater than 1 is considered as containing mixed emotion.

Quantitative data was collected and analysed in this paper. The anonymised posts and comments (text) in each category of the Google+ community of the July 2017 Intake were first collected by a computer programme written by a team member. The timeline of the text was also included in the gathered data. This cohort was chosen as its

students had completed their studies when the research project for this paper started. The text was then fed into Google’s Cloud-based Natural Language Processing tool to anonymously analyse the sentiment. Each post/comment on the G+ community is treated as a single document by Google’s NPL tool. Microsoft Power BI was then used to visualise the analysed data. The score and magnitude of posts and comments were analysed in terms of timeline and category. It should be noted that some posts were made by students in this cohort in Te Reo Māori, but at the time the research project was carried out, Google’s Natural Language Processing was not able to process Te Reo because a Te Reo Māori sentiment lexicon does not exist in the tool. Therefore, sentiments in these posts and comments were not analysed in this paper.

## Results and Analysis

### Overview of collected data

Table 4 shows the overview of the data collected in the G+ community. The messages on the online community are grouped into two main categories: initial posts, which are the first message in a thread of conversation, and comments that are the subsequent messages in those threads.

**Table 4: Overview of the collected data**

Number of students who either post initial posts or comment on the G+ community	314
Number of students commenting	<b>296</b>
Total number of initial posts	<b>2866</b>
Total number of comments	3630
Initial posts in English (The service can’t process Te Reo)	2844

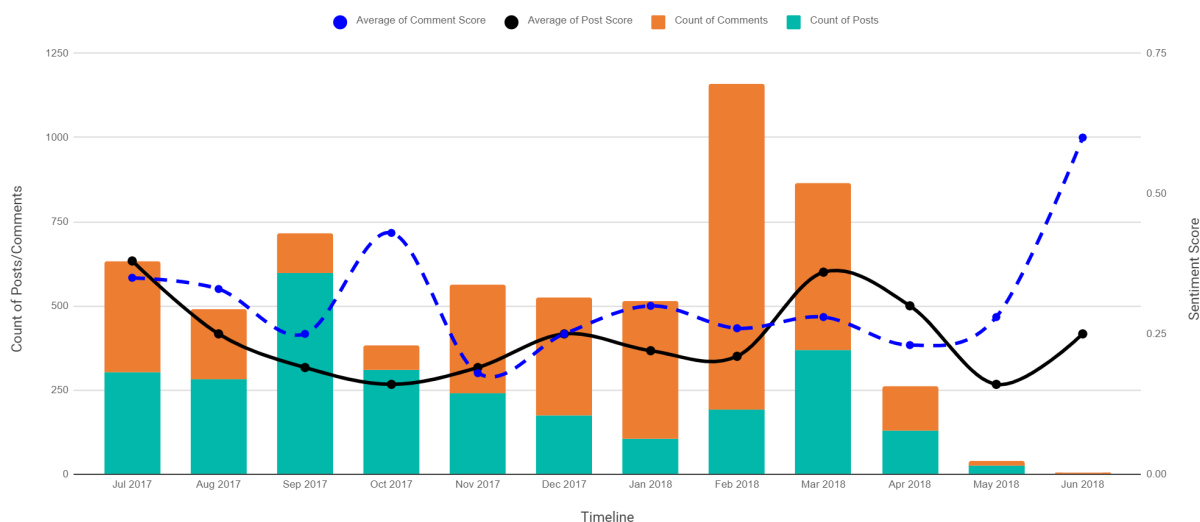
All students joined the G+ community. As shown in table 4, there was only a very small number of students (314 - 296 = 18) who posted only the first messages in the threads of conversation. The majority of students interacted with other students by commenting on an initial post of a fellow student or replying to a comment on their initial post. While staff members were very active in replying to questions from students on the G+ community or commenting on their messages, a significant majority of the initial posts and comments were from students.

### Timeline-based Analysis of Online Sentiment

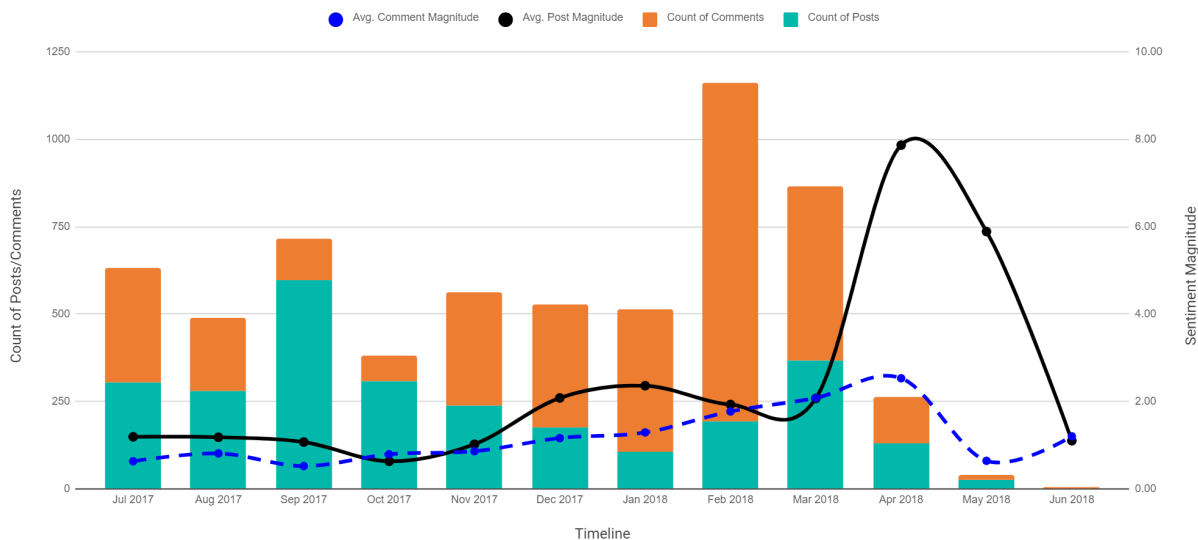
Figures 1 and 2 show the variation of students’ online sentiment throughout the programme (score in Figure 1 and magnitude in Figure 2) based on the posts and comments up to 12 months since the start of the programme. The first 16 weeks from July to November is when students had weekly face-to-face workshops. After that, they started the 16-week online learning from November to March including a two-week break during December. Students still contribute to online discussion after they finish their studies up to June. However, the numbers of posts and comments in May and June 2018 (two to three months after the official end of the programme) were very small, therefore, the average sentiment scores are not accurate anymore. There are some posts on the G+ community that contain only a hashtag or an URL. However, their impact on the sentiment analysis is insignificant as the graphs shown in both figures 1 and 2 are almost the same when they are removed.

As shown in figure 1 and 2, the average sentiment score of the posts is higher than the average sentiment score of the comments during the first month of the programme, then it is lower for the rest of the first 16 weeks. The average sentiment score of the posts is also higher than the average sentiment score of the comments during the first month of the online courses, then it is lower for the next 10 weeks before getting higher again in the last two weeks of these papers. Figures 1 and 2 also show a decrease in the number of comments per month during the first 16 weeks. This could be attributed to the face-to-face nature of the first two courses which reduce the need for online interaction. The number of comments is usually less than the number of initial posts during these papers. During the first half of the programme, students attended weekly face-to-face sessions and shared the outcome of their activities during the sessions on G+. Students were encouraged to participate in G+ so that they would be confident with interacting on the platform (or social media in general). This would help them to confidently seek help or exchange ideas during the second half of the programme when they did the online courses and did not see their peers every week. As the students saw their peers and facilitators every week during the first 16 weeks, the need to express their sentiments on the G+ community was very likely to be lower when compared with the second 16 weeks.

Most of these posts are in the categories ‘In Class Tasks’ and ‘Student Reflection’ for the first 16 weeks of the programme.



**Figure 1: Monthly average sentiment score of initial posts and comments during the programme.**

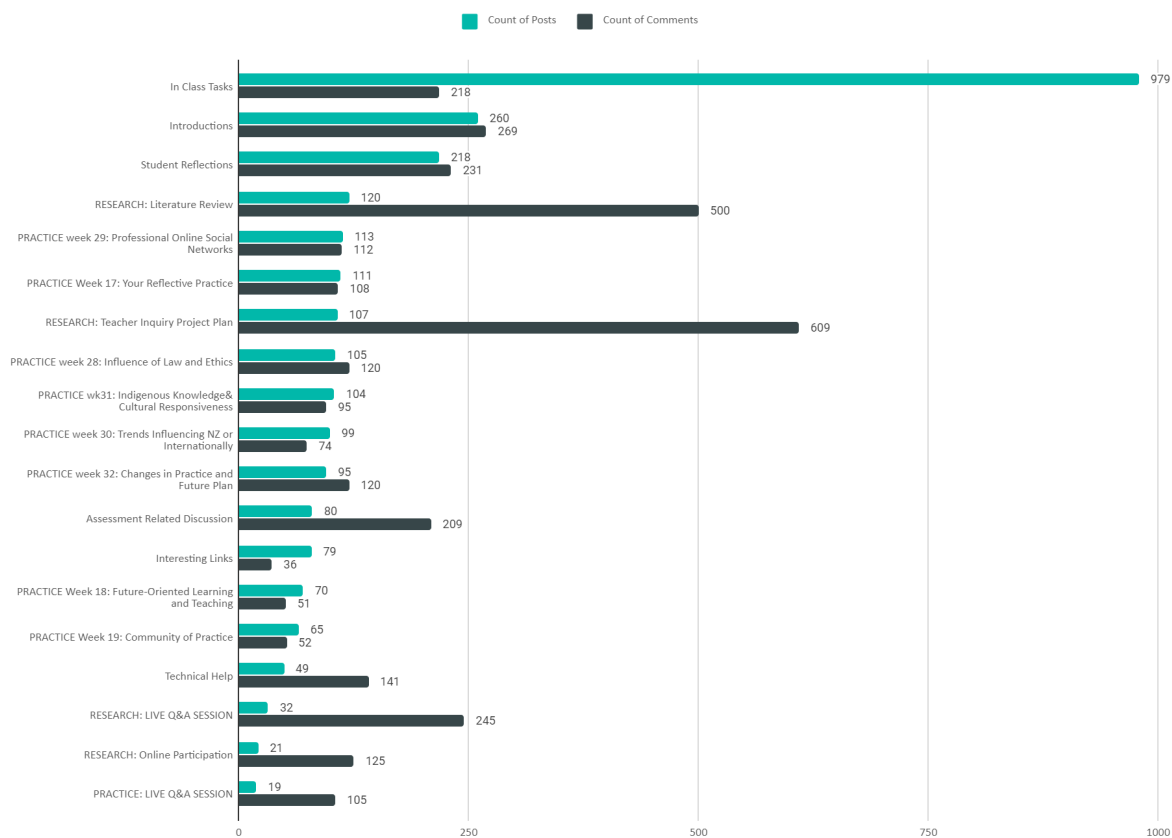


**Figure 2: Monthly average sentiment magnitude of initial posts and comments per month during the programme.**

For the second part of the programme, the number of first posts per month reduces gradually from the first month to the third month, which was the Summer holiday, and keeps increasing in the last two months. The high number of initial posts could be attributed to the sharing of blog posts in the PRACTICE paper.

Figure 3 shows the count of posts and comments for each category in the G+ community. The ratio between the number of comments and the number of initial posts increased in the second half of the programme. In particular the numbers of comments for categories “RESEARCH: Literature Review” and “RESEARCH: Teacher Inquiry project plan” are very high. The monthly average values for both sentiment score and magnitude also increased during the second half of the programme.

Closer investigations of the curves show that the average sentiment score of the comments is highest in October 2017, while during March 2018, the average sentiment score for the initial posts has the highest value. Most of the online conversations during October 2017, which is in the first 16 weeks of the programme, are in the sections ‘In Class Tasks’ and ‘Interesting Links’, and the topics for the four weeks of the month are ‘Design Practice’, ‘Game and Gamification’, ‘Design Thinking’, and ‘Teacher Inquiry’. The high value of sentiment score of comments in October could be attributed to the topics taught in that month. For the month of March 2018, students worked on the last four reflections required in the PRACTICE course. The topics for these four reflections are ‘Professional Online Social Networks’, ‘Trends Influencing NZ or Internationally’, ‘Indigenous Knowledge & Cultural Responsiveness’, and ‘Changes in Practice and Future Plan’. The high sentiment score of the initial posts may be attributed to the relevance of the reflective topics to students’ practice and that they had nearly completed the programme.

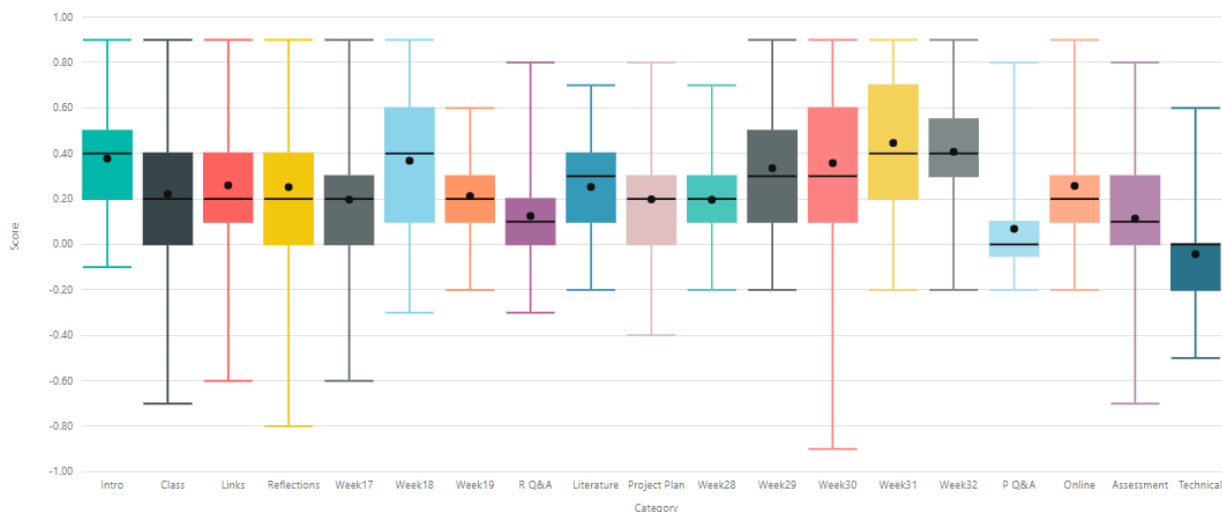


**Figure 3: Total count of posts and comments for each category in G+ community.**

### Categories-bases Analysis of Online Sentiment

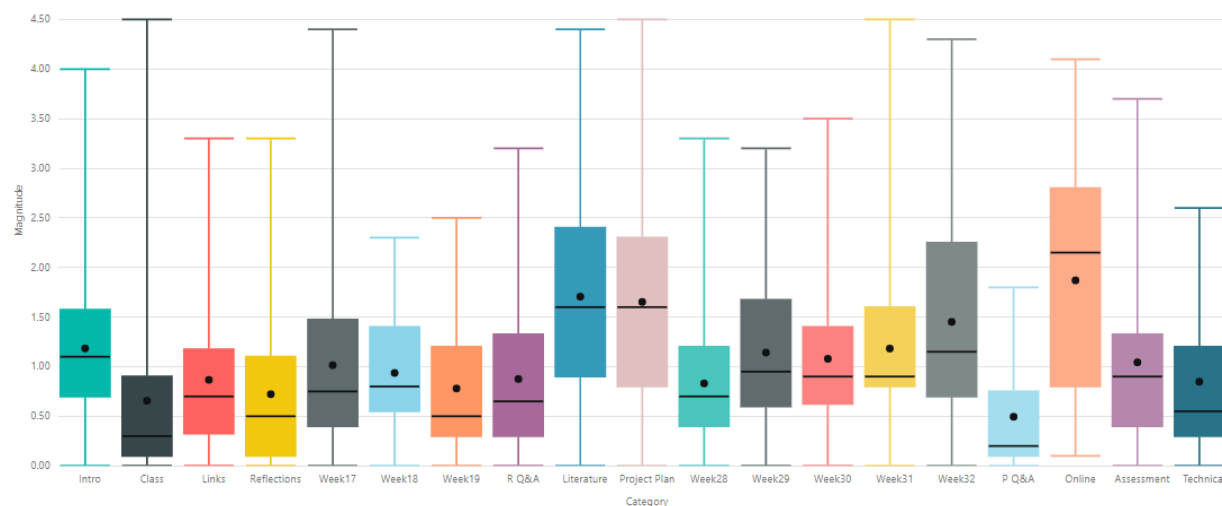
The previous section looked at the variation of sentiment in students’ initial posts and comments with time. The next two figures look at the distribution of sentiment score and magnitude in each category. Figure 4 shows the box plot of sentiment score for both initial posts and comments in each category of the G+ community. Different colours are assigned by the data visualisation tool, which is Microsoft Power BI, to box plots of different categories. The box plot for each category shows not only the distribution of the scores but also their average and median values. The combination of these three factors would help to show a clear picture of what sentiment scores look like in each category. An analysis based on only one of them is likely to result in an inaccurate conclusion. Figure 4 shows that while most of the initial posts and comments in most of the categories had a positive sentiment score, in several categories such as Technical Help, Assessment Related Discussion, PRACTICE: LIVE Q&A SESSION, RESEARCH: LIVE Q&A SESSION, the sentiment score of a majority of the posts and comments is within the range from -0.25 to +0.25. Therefore, the values of the sentiment magnitude should be looked at to determine whether the respective sentiment is actually neutral or if the text contains mixed emotions.





**Figure 4: Box plot for sentiment score of each category in the G+ community, the black dot in each box indicates the average value and the black horizontal line in each box shows the median value.**

Figure 5 shows the box plot of sentiment magnitude for those categories. Similarly to figure 4, the box plot for each category shows not only the distribution of the magnitudes, but also their average value and median values. The variation of box plot colours has the same nature. Figure 5 shows that, for many categories, the majority of the posts and comments have a sentiment magnitude greater than 1. This means that the amount of emotions in the posts and comments are not insignificant. Sentiment magnitude in several categories such as RESEARCH: Literature Review, RESEARCH: Teacher Inquiry Project Plan, and RESEARCH: Online Participation, and PRACTICE Week 32: Change in Practice and Future Plan, has a wide range of variation. These categories belong to the two online courses in the second part of the program.

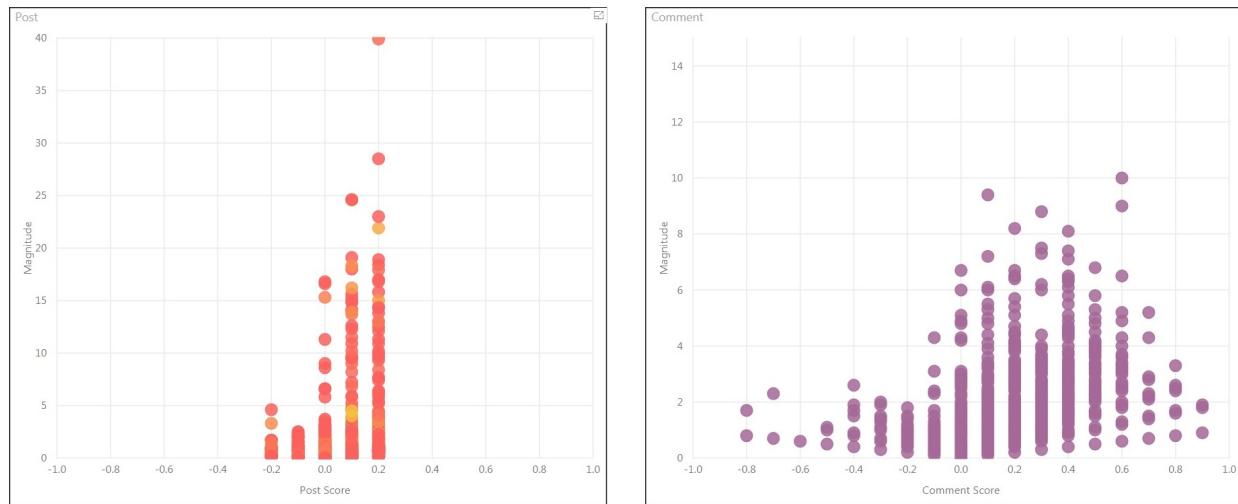


**Figure 5: Box plot for sentiment magnitude of each category in the G+ community, the block dot in each box indicates the average value and the black horizontal line in each box shows the median value.**

Figures 4 and 5 show that while the sentiment scores, which indicate whether the sentiment is negative or positive, of the posts and comments are not much different between the face-to-face courses (the first 16 weeks of the

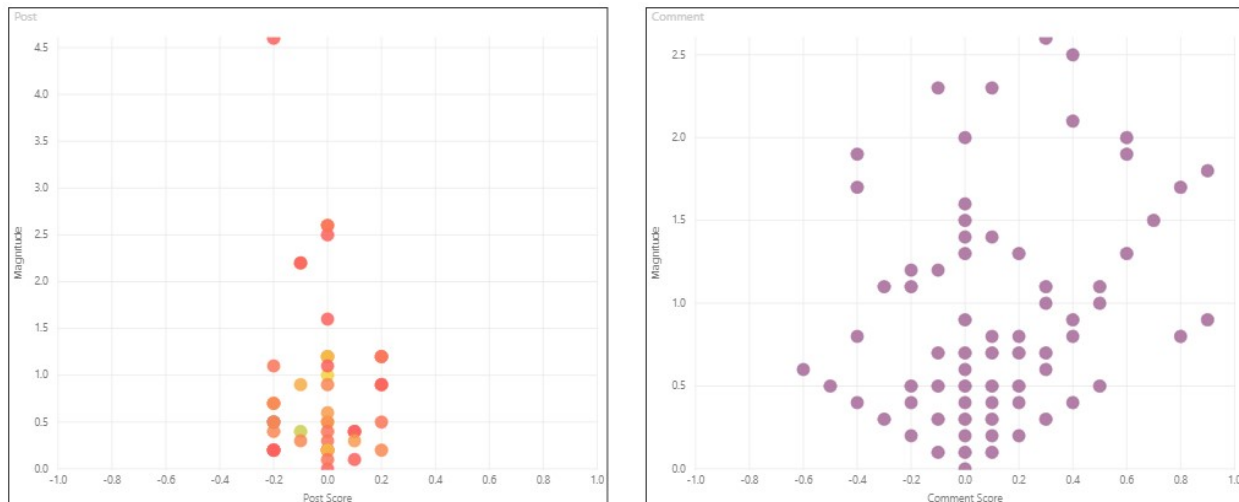
programme) and the online courses (the second 16 weeks of the programme), the sentiment magnitude, which is the amount of emotion, is more significant in the posts and comments in several categories of the online courses.

As figure 4 shows many sentiment scores are in the near-zero range. A closer look at these posts and comments is necessary. Figure 6 presents the sentiment magnitude of the posts in the G+ community that have sentiment score in the range from -0.25 to +0.25. Different colours in the left graph denote different numbers of subsequent comments for each initial post. The score and magnitude of comments in the respective thread of conversation are also shown in the graph on the right. The left graph shows that a large number of those posts contain mixed emotions, not being sentimentally neutral. From the graph on the right, while initial posts might be neutral, the sentiment score of the following comments can be clearly either negative or positive. The amount of emotion, which is expressed by sentiment magnitude, is also quite significant in them.

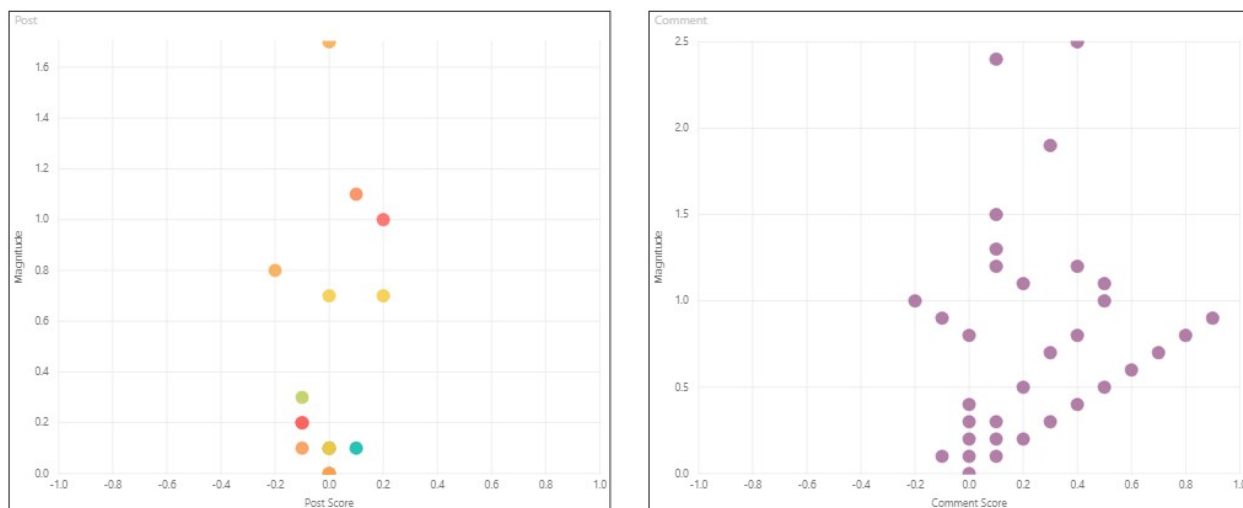


**Figure 6: Score and magnitude of posts in the G+ community with low sentiment score and their respective comments.**

In addition to figure 6, the categories ‘Technical Help’ and ‘PRACTICE: LIVE Q&A SESSION’ are chosen to be further investigated as it has been shown in figure 4 that almost all the initial posts and following comments in each of them have a sentiment score in the range from -0.25 to +0.25. Similar to figure 6, figures 7 and 8 show the distribution of sentiment score, which is in the range near zero, and the corresponding magnitude in the graph on the left. Sentimental score and magnitude of comments, which are in the respective threads of conversation, are in the right graph.



**Figure 7: Analysis of the posts with the sentimental score in the range [-0.25,0.25] in category Technical Help.**



**Figure 8: Analysis of the posts with the sentimental score in the range [-0.25,0.25] in category PRACTICE: LIVE Q&A SESSION.**

From figure 4, there are not many posts with a very negative sentiment score although category Technical Help is where students might express frustration due to technical issues. This could be attributed to the fact that this category is for the first 16 weeks of the programme when students can ask for assistance from the facilitators in weekly face-to-face interactions. Sentiment score and magnitude in the left graph of figure 7 suggest that most of the posts with low sentiment score are actually neutral, there is a small number of posts with mixed emotion. For the following comments, there are similar proportions of them with positive sentiment, negative sentiment, neutral, and containing mixed emotions. In figure 8, there are only two initial posts that contain mixed emotions, the rest of them are neutral. For the following comments, there are almost similar proportions of them with positive sentiment, neutral, or containing mixed emotions.

This paper also looks at categories, in which the students’ sentiment was observed to be negative or very negative by the facilitators, such as ‘Student Reflection’ for the first 16 weeks. For the second 16 weeks, the categories were

‘RESEARCH: Literature Review’ and ‘PRACTICE week 19: Community of Practice’ as these assessments were more demanding than the other ones. As there were very large numbers of posts and comments in those categories, the sentiment in the messages could not be reasonably accurately evaluated by just the observations. The actual sentiment in the three categories is analysed in Table 5. The table shows the average and median values of the sentiment scores and magnitude of the initial posts and comments.

**Table 5: Average and Median of Score and Magnitude in Categories Expected to Be Sentimentally Negative**

<b>Category</b> <i>Statistics</i>	<b>Student Reflection</b>	<b>RESEARCH: Literature Review</b>	<b>PRACTICE Week 19: Community of Practice</b>
<i>Average score of initial posts</i>	0.25	0.25	0.21
<i>Median score of initial posts</i>	0.20	0.30	0.20
<i>Average score of comments</i>	0.44	0.33	0.27
<i>Median score of comments</i>	0.45	0.30	0.30
<i>Average magnitude of initial posts</i>	0.87	1.75	1.58
<i>Median magnitude of initial posts</i>	0.50	1.60	0.70
<i>Average magnitude of comments</i>	1.13	1.33	2.01
<i>Median magnitude of comments</i>	0.90	0.9	1.40

Table 5 shows that the sentiment in the initial posts in category ‘Student Reflection’ is truly neutral while the emotion in the comments is slightly mixed. In both categories ‘RESEARCH: Literature Review’ and ‘PRACTICE Week 19: Community of Practice’, the initial posts contain mixed emotion while the sentiment in the comments is slightly positive. Therefore, the sentiment in students’ interactions in those three categories is better than expected. Our hunch developed during the delivery of the course was that the sentiment in these categories was negative as indicated by some of the initial posts and comments. With the data collected, we are now able to find out that sentiment in those categories was overall positive which was contrary to our initial hunch.

The results presented in figures 1, 2, 4, 5 and Table 5 have shown that the sentiment in students’ online interaction during the two online courses (the second 16 weeks) is more positive than in the face-to-face courses (the first 16 weeks). The posts and comments also contain more emotion during those two courses. Figure 3 also shows that online interaction during online courses is also higher. These could be attributed to the weekly face-to-face contact

with fellow students and facilitators during the first 16 weeks of the programme which provides a traditional communication channel for students.

## Discussion

As mentioned in the literature review section of this paper, most studies in sentiment analysis in education (Kim & Calvo, 2010; Nitin, Swapna & Shankararaman, 2015; Balahadia, Fernando & Juanatas, 2016; Dhanalakshmi & Bino, & M. Saravanan, 2016; Rani & Kumar, 2017; Aung, K & Myo, 2017) look at student feedback data. However, Cleveland-Innes and Campbell (2012)'s research shows evidence of emotions present in online learning, especially in an online community. This study, therefore, investigates the online sentiments of students in their posted texts on the G+ community of learning with the considerably high volume of numbers of initial posts and comments, 2866 and 3630 respectively. Another aspect that this paper might contribute to the research in this area is that while studies in sentiment analysis in education mostly focus on specific machine learning classification techniques such as that of Sivakumar and Reddy (2017); Ozturk, Cicek and Ergul (2017) and Altrabsheh, Cocca and Fallahkhair (2014), this research uses the available cloud-based tool, the Natural Language Processing on Google cloud platform to observe the online sentiments via textual posts in terms of the overall sentiments, sentiments according to timeline and sentiments in the G+ community categories. The tool's characteristic of providing two outcomes for sentiment analysis, which are sentiment score and magnitude, has given the research extra benefits to analyse both the polarity of the sentiment and amount of emotion as other cloud-based tools such as IBM Watson Natural Language Understanding and Microsoft Azure Text Analytics API can only give a single output that indicates the polarity of the sentiment (Vo & Pham, 2019). Google Cloud Platform also performs better than a wide range of other sentiment analysis tools in the comparison discussed in Vasilescu et. al (2019).

Another benefit of this research is the use of Microsoft Power BI as the tool to analyse and visualise the outputs from the Google Cloud Platform. This tool's ability to quickly show and analyse the whole range of data is particularly beneficial when ranges of score and magnitude are quite wide or narrow. For example, when the averages of the score for categories are quite similar, the plotting and analysing of the whole range of score provide more adequate information about the distribution of sentiment in each category. Also if the scores are neutral, it shows the whole range of sentiment magnitude. This allowed researchers to create an overview of mixed emotions.

The observation of the interaction of students on the online community shows that there is an increase in the number of initial posts and comments in the second half of the programme when the weekly face-to-face sessions were not available anymore (Pham, Vo, Lindsay, Li, Pashna, Baker, Han, Rowley, 2019). The increase is also indicated by the ratio between the number of comments and the number of initial posts in each category as shown in figure 3 of the previous section. There is also an increase in the emotion expressed when students begin the online courses, and this is backed up by Cleveland-Innes and Campbell (2012) when talking about transitioning to a new learning environment. The high volume of contributions and interactions on the online community has made it hard for instructors to know students' sentiment during their learning to provide timely help for this intake and improve the learning support for the next cohort.

The data analysis has shown that the sentiment, in general, is more positive than expected. The analysis in each category of the G+ community also shows that the sentiment is more positive than expected in categories for the assessments of the two online courses in the second half of the programme. As many students in every intake of the programme had not done online learning before, instructors, who delivered the online courses often received student complaints and expressions of frustration. Several different channels for learning support, which are both inside the G+ community (such as weekly Live Chat sessions, timely and regular responses/comments from the instructors) and outside it (such as personal online meetings, weekly webinars), have been provided to students to help them complete their studies on time. However, the impact of those supports has only been measured by qualitative data

such as student feedback at the end of each half of the programme and quantitative data such as student assessment results. Since Cleveland-Innes and Campbell (2012) indicated that texts from interaction in online community postings, compared to the single-person response to questions in a survey, demonstrated more “numerous, complex, and interrelated emotional tone” (p.282), this extensive quantitative data analysis with this cloud-based tool has provided a better understanding of students’ emotion about their learning through their online interactions during the programme. The analysis also shows that some categories in the G+ community for the first sixteen weeks, which were quite likely to have negative sentiments such as Technical Help and Student Reflection, also contain more positive sentiment than expected.

The online sentiment was more positive at the beginning of the first 16 weeks when the programme started (this was consistent with what was observed in the face-to-face sessions). This positivity could be attributed to the enthusiasm of the students, who were full-time registered teachers, to come back to formal studies. Then the sentiment got less positive toward the end of the first half and the start of the second half. The sentiment was more positive again during several weeks near the end of the second 16 weeks. Once the students got the momentum in their learning and were more fluent in digital skills, their sentiment was more positive. Quantitative data about the timeline of student sentiment during the programme could help to provide timely support for the subsequent intakes.

## **Conclusion**

This paper is about an experience in applying a readily available cloud-based Natural Language Processing Tool in an educational setting to determine its applicability in education. The research investigated student sentiment in online postings, using predetermined categories, over periods of time during a postgraduate course.

Overall the Google Cloud Natural Language API demonstrated that when sentiment is high, students are expected to cope well with their learning in contrast to the low sentiment which indicated that students might not be coping well with their learning and more support may need to be provided. In this programme, the general increase from Nov 2017 onwards in online sentiment arose because this was when the online courses began and included online interaction as assessment.

There were several possible reasons why students’ online sentiment is getting more positive in either initial posts or comments. The first of these tended to be an increase in confidence in the online platform itself over time, while the second was the use of provocations or interesting topics to discuss as these tended to create more conversations. Thirdly, when there were not weekly face-to-face sessions and students discussed the assessments of the online courses which awarded small marks for online participations. However, the sentiment in initial posts and that in comments were not always getting more positive (or negative) at the same time.

One of the limitations of this research was that the natural language processing tool was unavailable in Te Reo which meant some students were excluded from data analysis meaning the complete cohort sentiment cannot be fully comprehended. This might be an opportunity in future for an API to be offered in Te Reo or an indigenous language. In addition to that, the specific methods used by Google Cloud Platform to analyse the polarity and amount of sentiment are not revealed by the company. Therefore, the analysis of the tool’s outputs is limited.

Although we would like to explore the use of cloud-based Natural Language Processing Tool in MeWe, which is replacing our social media for G+, a foreseeable potential limitation of social media (Facebook / MeWe) is the inability to offer categories like G+ currently offers. Therefore another approach to data collection using this natural language processing tool may be needed such as the use of hashtags to replace categories. The limitation of hashtags may be students inconsistently using hashtags and also using pre-determined keywords spelt correctly.

In future, sentiment analysis may be an opportunity to follow focus groups of struggling and high performing students over time and analyse how their sentiment varies during the programme at specific points indicating intervention points. Sentiment analysis allowed us to verify hunches using data regarding sentiment as we could investigate the extent to which sentiment was experienced by a large group of students rather than a few. As shown in the analysis, at times our hunches were proven to be incorrect by the data. Sometimes in a course, there may be a student who is very vocal either negative or positive that can be influential. We would like to gather qualitative data to investigate further into the quantitative data we have gathered to cater better to student needs.

The research question was ‘How does a natural language processing tool help analyse student online sentiment in a postgraduate program?’. Google Cloud Natural Language API is significant because it includes score and magnitude therefore researchers can investigate not only whether the sentiment is positive or negative but also the amount of emotional expression in online contributions. This may allow educators to tailor programmes to meet the needs or interests of students. Educationalists or course designers could initially use sentiment analysis to determine high and low sentiment scores in terms of timeline or discussion topics and use student surveys to seek an understanding of phenomena. This tool should not be used by itself, but in conjunction with student feedback and teacher observation.

## References

- Altrabsheh, N., Cocea, M., & Fallahkhair, S. (2014). Sentiment analysis: towards a tool for analysing real-time students feedback. In *Proceedings of IEEE International Conference on Tools with Artificial Intelligence, ICTAI*, pp.419–423. <https://doi.org/10.1109/ICTAI.2014.70>
- Aung, K., & Myo, N. (2017). Sentiment analysis of students' comment using lexicon based approach. In *Proceedings of IEEE/ACIS 16th International Conference on Computer and Information Science (ICIS)*, pp. 149-154, Wuhan. doi: 10.1109/ICIS.2017.7959985
- Balahadia, F., Fernando, M., & Juanatas, I. (2016). Teacher's performance evaluation tool using opinion mining with sentiment analysis, In *Proceedings of IEEE Region 10 Symposium (TENSYMP)*, pp. 95-98, Bali. doi: 10.1109/TENCONSpring.2016.7519384.
- Cleveland-Innes, M., & Campbell, P. (2012). *Emotional presence, learning, and the online learning environment. The International Review of Research in Open and Distance Learning 13(4)*.
- Colace, F., De Santo, M., & Greco, L. (2014). SAFE:a sentiment analysis framework for E-Learning. *International Journal Of Emerging Technologies In Learning (IJET)*, 9(6), 37. doi: 10.3991/ijet.v9i6.4110
- Dhanalakshmi, V., Bino, D. M., & Saravanan, A. (2016). Opinion mining from student feedback data using supervised learning algorithms. In *Proceedings of 3rd MEC International Conference on Big Data and Smart City (ICBDSC)*, pp. 1-5, Muscat. doi: 10.1109/ICBDSC.2016.7460390
- Google. (2020a). Natural Language. Retrieved at <https://cloud.google.com/natural-language>.
- Google. (2020b). Natural Language API Basics. Retrieved at <https://cloud.google.com/natural-language/docs/basics#sentiment-analysis-values>
- Kim, S., & Calvo, R. (2010). Sentiment analysis in student experiences of learning. In R.Baker, A. Merceron, P.Pavlik Jr. (Eds.), *Proceedings of Educational Data Mining 2010, The 3rd International Conference on Educational Data Mining* (pp.111-120.). Pittsburgh,, USA.
- Liu, B. (2012). *Sentiment analysis and opinion mining*. San Rafael, Calif.: Morgan & Claypool.
- Mäntylä, M., Graziotin, D., & Kuutila, M. (2018). The evolution of sentiment analysis—A review of research topics, venues, and top cited papers. *Computer Science Review*, 27, 16-32. doi: 10.1016/j.cosrev.2017.10.002
- Nitin, G. I., Swapna, G., & Shankararaman, V. (2015). Analyzing educational comments for topics and sentiments: A text analytics approach. *Proceedings - Frontiers in Education Conference, FIE*. <https://doi.org/10.1109/FIE.2015.7344296>
- Ozturk, Z. K., Cicek, Z.E., & Ergul, Z. (2017). Sentiment analysis: an application to Anadolu University. *Acta Physica Polonica A*, 132 (3). 753-755.
- Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and trends in information retrieval*, 2(1-2), 1-135. doi: 10.1561/1500000011

- Rani, S., & Kumar, P. (2017). Sentiment analysis system to improve teaching and learning, *Computer, 50(5)*, 36-43, . doi:10.1109/MC.2017.133
- Sivakumar, M., & Reddy, U. S. (2017). Aspect based sentiment analysis of students' opinion using machine learning techniques. *Proceedings of International Conference on Inventive Computing and Informatics (ICICI)*, pp. 726-731, Coimbatore.
- Thelwall, M., (2016). Sentiment analysis. In L.Sloan, & A.Quan-Haase (Eds.) *The SAGE Handbook of social media research methods*(pp. 545-556), London: SAGE Publications Ltd. doi: 10.4135/9781473983847.
- Vasilescu, C., Suciu, G., & Pasat, A.(2019). A new method to help the human resources staff to find the right candidates, based on Deep Learning. In *Proceedings of The 15th International Scientific Conference eLearning and Software for Education*, (pp.240-246), Bucharest.
- Vo, D., & Pham, T. (2020) Sentiment Analysis in Education. In: Peters M., Heraud R. (eds) *Encyclopedia of Educational Innovation*. Springer, Singapore.