Human and Artificial Intelligence for the Society of the Future
European Distance and E-Learning Network (EDEN) Proceedings
2020 Annual Conference | Timisoara, 22-24 June, 2020
ISSN 2707-2819
doi: 10.38069/edenconf-2020-ac0016

A CONCEPTUAL FRAMEWORK FOR REAL-TIME EMOTIONAL-STATE MONITORING OF STUDENTS IN VLES TO IDENTIFY STUDENTS AT RISK

Lubna Alharbi, University of Liverpool, United Kingdom, University of Tabuk, Kingdom of Saudi Arabia, Floriana Grasso, Phil Jimmieson, University of Liverpool, United Kingdom

Abstract

Virtual Learning Environments (VLEs) feature rich textual data which lend itself naturally to the identification and monitoring of aspects of students' interactions. While reducing attrition and improving performance remain the primary objectives of learning analytics, we contend that student contributed text can be used to learn about emotions and other extra-rational features. This would help provide a response to the recent cries for help from the sector, seeking a system looking to address the worrying mental health crisis trends. This paper addresses these issues by discussing the necessary mechanisms within a conceptual framework which would sit in a VLE and capture emotional state changes in the students' interaction style or tone. For such a framework, the aim would be to help educators to carry out timely interventions when a potential cause of distress is identified. Experimental results on available datasets from education and psychology serve as a feasibility study for these tasks, and offer a perspective on the potential of the approach.

Introduction

Higher education is witnessing a mental health crisis. The largest mental health survey by the institute for Public Policy Research (Thorley, 2017) reported that the number of students who were diagnosed with a mental health condition had risen dramatically over the past 10 years. Students are at increased risk of dropping out because of the lack of support and treatment for mental health issues. HESA (https://www.hesa.ac.uk/) recently reported that almost 38,000 UK students are suffering from some form of psychological distress, with increasing levels of anxiety, loneliness, and thoughts of self-harm.

Current research has explored various Artificial Intelligence (AI) techniques in order to make the learning process more optimised to the behavioural states of students, and AI techniques for monitoring students online have been commonly used to identify behaviour patterns with the aim of increasing learning and retention rates (Avella, Kebritchi, Nunn, & Kanai, 2016), or in early prediction of at-risk students (Marbouti, Diefes-Dux, & Madhavan, 2016). The common assumption is that time spent on learning is related to academic performance, either positively (Fritz, 2011) or indeed negatively, as a predictor of some struggle (Lust, Elen, & Clarebout, 2013). So far however very few studies have paid sufficient attention to the analysis of the context around the learning activity, for example by using emotion and/or sentiment analysis or writing style features in addition to the student's educational history for better understanding and prediction of student engagement. We seek to explore the use of emotion and textual analysis to support the educators in understanding whether a situation needs their intervention. We concentrate in our investigation on fully online e-learning systems, and in particular systems where communication from the student's part is solely textual, in the belief that emotions can be detected in different on-line learning contexts (e.g. forum discussions, chat discussions) (Efklides & Volet, 2005). VLEs are rich text-based environments. Students communicate with the lecturer by asking questions, complaining, or seeking advice. This happens over time, during the course of an entire module, and possibly a programme of study. We therefore focus on recognising the features from the textual messages interchanged between students and their peers or teachers to help in monitoring the situation over time, and identify changes in those patterns.

This paper presents an analysis of the issue at hand, by describing a conceptual framework for incorporating an "emotional observer" into a VLE. We identify the tasks that need to be carried out, and we explore the feasibility of such tasks by presenting two experiments which tested each of the tasks on suitable datasets.

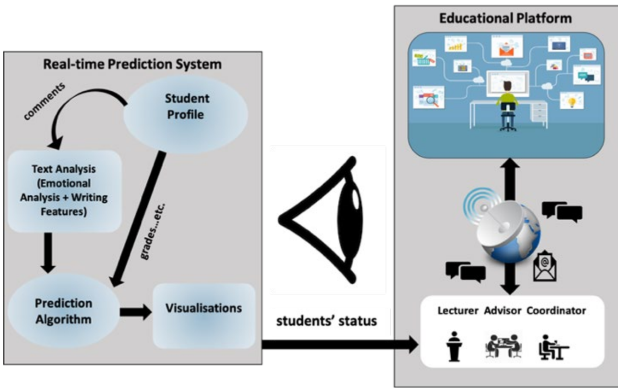

Figure 1. Structure of the "observer" system (Alharbi, Grasso, & Jimmieson, 2019)

Towards an Emotional Dashboard – Feasibility Experiments

The work presented in this paper is aimed at demonstrating the feasibility, as well as any technical challenges, of an "emotional observer" system, which would sit within a VLE and

ISSN 2707-2819

would be able to support educators in identifying potential situations at risk. Our ideal system is depicted in Figure 1, and would comprise a prediction system, which can decide to intervene, or simply flag, any observed interesting emotional communications. In our design, such a system should be able to analyse online communications happening in a virtual classroom, and identify emotionally loaded text, as well as flag to the educators the situation that such text is associated with. In order to do so, such a system would need to be able not only to perform emotion analysis of a piece of text, but also to identify other factors which could be a sign worthy of attention, for example a change in the writing style of a student, or whether a change can be detected in the correspondence of a specific topic of discussion.

Emotions have been extensively studied in behavioural sciences and psychology, where the two main approaches are the categorical, (e.g. Ekman & Friesen, 1971), which considers only a limited number of “basic emotions” (typically: anger, disgust, fear, happiness, sadness, and surprise) and the dimensional, (e.g. Russell, 1980) which places emotions on a two-dimensional space, with “Arousal” on one level, and “Valence”, positive or negative, on the other. In education, research has demonstrated the significance of sentiment/emotions within the e-learning processes by highlighting their impact on academic achievement (Artino, 2012; Pekrun, 2006), and the role they play in engagement (D’Errico, Paciello, & Cerniglia, 2016). Advances in computational linguistics reassure us that it is indeed possible to extract emotions, opinions and sentiments from text (Strapparava & Mihalcea, 2008).

Our framework proposes to combine emotion analysis with the analysis of style, a task well studied in stylometric analysis (Abbasi & Chen, 2008), typically for the purpose of identification of authorship (Zheng et al., 2006), and as a necessary step in the construction of an observer system as described above, we tested its feasibility, not only from a purely technical point of view, but also as to the practicality of such a solution. We therefore conducted a set of experiments, using well established data analysis techniques, in order to ascertain whether we can in fact find emotionally loaded text in classroom interactions, and up to which extent discovering emotionally loaded text is useful to predict user performances (task 1). We also investigated which emotion is more likely to be a better predictor, and whether we can think longitudinally and identify ”trends” in the emotional state and/or writing style of an individual student, with respect to other students in the same classroom, or with respect to previous interactions of the same student (task 2). The latter point would confirm the potential for an observer system to determine a sort of “emotional baseline” for a student, or a class, or indeed a topic, and flag any significant deviation from this baseline as something potentially worthy of attention by a human educator.

The greatest challenge in attempting our experimental studies was to identify suitable datasets, and while task 1 could be performed on datasets coming from educational resources, for task 2 we used material from a different setting. The dataset description, as well as the details of each experiment, are explained in what follows.

First Experiment: the Stanford MOOC dataset

In this experiment, we focused on pre-processing and analysing students' data, and developing a model to predict students' performance using data mining (Kantardzic, 2011). The objective of this analysis is to understand any general relationships between different student's characteristics, including the emotions features, and the prediction of student final assessments. The experiment consisted of two phases, where a prediction of the final grade of a student is attempted without and then with the emotions analysis features. The experiment concludes with the selection of which features, according to ranking, and which emotion feature would provide the best performance.

The dataset used in this phase is excerpted from Stanford MOOCs dataset (http://datastage.stanford.edu) by the Center for Advanced Research through Online Learning (CAROL), made available for use by researchers and instructors. The dataset consists of a number of tables containing various anonymised information from the MOOCs activities.

From all the items in the dataset, we extracted 229 records, related to students who could be followed through all data files. Among the tables and views provided, we concentrated on ActivityGrade containing grades, right/wrong answers and student's choice, and submission times; FinalGrade containing the grades computed at the end of the course; EdxForum.contents containing text entries to the discussion forums; and Alldata containing the name of the course, and other attributes such as SessionLength (sec). We classified the final grades into four classes: Fail, <50%; Pass, 50-59%; Merit, 60-69% and Distinction, >=70%.

To analyse the dataset, we used WEKA (Waikato Environment for Knowledge Analysis) version 3.8.2 (Hall et al., 2009), as it allows for an easy delivery of an analysis report and a prediction model. The WEKA classifier used in this study is a Naive Bayes Classifier (NBC), as recommended elsewhere (Mueen et al., 2016; Devasia et al., 2016), in the spirit not so much of finding the best possible prediction, but to simply demonstrate the feasibility of the task using off-the-shelf tools.

ISSN 2707-2819

Phase 1: Testing the Dataset without Emotions Features

For the first phase in our experiment, we selected the following features: (Percent Grade, SessionLength, NumEventsInSession, Module Type, and FinalGrade) and we used them to train an NBC, using cross-validation with k-fold 10. The accuracy result for using an NBC to predict a student FinalGrade was 80% (with 0.75 precision, 0.80 recall). We have therefore a benchmark to be able to determine whether adding emotional traits will give us any improvement.

Phase 2: Testing the Dataset with Emotions/Sentiment attributes

In this phase we wanted to ascertain whether adding emotional features extracted from the students' textual comments would improve the classification. First, we extracted all textual comments related to the 229 records identified in the first phase. This gave us 72 instances containing textual comments made by the students. Three classes were tested: Fail, Pass and Distinction as there were no instances for class Merit. The accuracy result for using NBC to predict a student FinalGrade for the 72 instances was 90.2%, with 0.90 recall.

Then in this phase to extract the emotion features, we employed a sentiment and emotion analysis tool, Synesketch (Krcadinac et al., 2013) for a sentence analysis of the comments and threads, and this extracted seven main emotions: *Anger, Disgust, Fear, Happiness, Sadness, Surprise and Valence*. Then we trained a Naive Bayes Classifier (NBC) using these emotional features plus the ones used in the previous phase. Again, we applied a 10-fold cross validation approach to assess the performance of the models. Including emotions/sentiment features gave us a better classification accuracy of 86.1% overall, with 0.86 recall.

Most Influential Features

Once established that class interaction textual elements do indeed contain emotional text, and that these could be used to predict the final grade, it is interesting to determine whether there is a specific emotion that is more *useful* than others to this task, and in general which features should be selected to maximise the results (Blum & Langley, 1997). Running a Features Selection process on our dataset suggested that one attribute, Session Length, has the highest correlation with the output class. It also suggested a host of attributes with some modest correlation, SessionLength(sec), NumEventsInSession, percent grade, Sadness, Happiness, Anger, and Up count. If we set our cut-off for relevant attributes equal to 0.1, then the remaining attributes (i.e., Disgust, Fear, Surprise, Valence, Down count) could possibly be removed. The result of the confusion matrix after removing the non-relevant attributes is 90.27%, with 0.90 recall. This is 4% higher than the one with all the emotion features included.

Second Experiment: The Motivational Interviews Corpus

That emotional features are present in students' exchanges, and that they can be used to a certain degree to predict the final outcome, is perhaps not surprising, and is not a novel result. Our main objective though was to conjecture a system where this information can be used in real time to inform the educators about possible areas of concern. In order for this to happen though, it is not sufficient to determine which student is exhibiting which specific feature, if there is no sense of what is "normal" for that particular student. Lacking any other visual cue which can be acquired by physical interactions, it is not easy to extrapolate whether a particular writing style, or a particular attitude in writing is a peculiarity of a student, or is a sign of a change in the emotional status, or ultimately of distress.

When sensors and physical traits are used to measure individuals, it is typical to create a baseline for vital characteristics, the change in which can be a sign of an underlying cause. Can this be done with textual traits too? To be able to understand whether this was conceivable, we needed to test it on a suitable dataset, and the Stanford MOOC used for the first experiment, or indeed the educational datasets we have been able to access opensource, are not ideal, as they concentrate on parametric values, and very rarely provide access to students' communications, and communication by the same student over time.

In this experiment we turned to the field of motivational interviewing (Rollnick & Miller, 1995) and we created a dataset by using a corpus of transcripts of dialogues, tracking the same individual over time, this providing us with the opportunity to test the feasibility of our approach, to be translated to the classroom environment. The corpus (Counselling and Psychotherapy Transcripts, Client Narratives, and Reference Works, 2008) consists of a searchable collection containing real transcripts of counselling and therapy sessions. The database contains more than 2,000 session transcripts, 44,000 pages of client narratives, and 25,000 pages of secondary reference material. For each transcript, which are anonymised but can track the same patient over a number of counselling sessions, the source provides information on age, gender, marital status, any symptoms or condition, as well as some general information on the therapist, such as gender and level of experience. Out of this corpus, we created a dataset, extracting data related to 76 patients. For each patient, there is a variable number of sessions. Each session is labelled with a title. For the purpose of this experiment, we only considered the patients' turns, we discarded the counsellor comments, and any other comments by the scribe (e.g. annotating pauses etc.).

ISSN 2707-2819

We performed a word level classification to the patient text in each session, by using the algorithm in Whissell (1996), using an *emotion clock*. We extracted around 12 emotion types using the clock: Peaceful, friendly, admiring, cheerful, brave, alert, alarmed, furious, sad, distant, depressed, indifferent. We used WordNet (Miller, 1998) and *dictionary.com* for capturing the lexicon and synonyms.

In addition to the emotional status, we wanted to test whether it would be useful to extract some characteristics of the text, which would help form the baseline for each individual, and some notion of "cohort", which would help form the baseline for a "classroom". For the first task, we extracted two features: Part of Speech (POS) and vocabulary richness of the text, by following the technique in Suh (2016) who maintains that changes in these parameters is an indication that the writer has changed their usual writing style. For POS we calculate the total number of complete sentences normalised by the total number of sentences in the session. To find the vocabulary richness of the text, we calculate the total number of unique words, normalised by the total number of words.

For the second task, we wanted to aggregate a number of client sessions in order to consider them as part of the same "cohort". The dataset provides us with a "topic" for each session, which is contained in the session title. We extracted a number of patients for which the topic of each session would align. These correspond to seven main topics: Behaviour, Development, Relationships, Culture, Ability, Personality, and Health.

Once all the machinery was in place, this gave us the opportunity to look at a set of patient sessions, and to use them to simulate the behaviour of a classroom. The working hypothesis is that each "topic" of discussion is the analogue of a topic in the classroom, and that each of the patients will react to different topics in their own different way. The observer system will be able to perform an emotion/sentiment analysis, a POS analysis, and the extraction of a specific topic, and decide if there is a situation to flag.

Row Labels	Max of Postagging	Max. of Richness
Ability	**0.90797546**	**0.614583333**
Admire	**0.492753623**	**0.478079332**
Alert	**0.642384106**	**0.487323944**
Brave	**0.666666667**	**0.545112782**
Cheerful	**0.618181818**	**0.451476793**
Friend	**0.83203125**	**0.41560219**
Furious	**0.61682243**	**0.467836257**
Not Available	**0**	**0.611111111**
Peace	**0.90797546**	**0.614583333**
Sad	**0.8**	**0.604**
client004_a	0.411255411	0.334792123
client004_i	0.450715421	0.200187091
client018_d	0.668639053	0.292576419
Client032_l	0.451428571	0.287434161
Client112_b	0.595744681	0.397034596
Client123_c	0.366666667	0.570512821
client212_c	0.366666667	0.397427653

[illegible]	[illegible]	[illegible]
client423_i	0.303191489	0.294806145
Culture	**0.790598291**	**0.469333333**
Admire	**0.657894737**	**0.469333333**
Friend	**0.790598291**	**0.302812071**
Peace	**0.628571429**	**0.427258806**
Development	**0.823529412**	**0.656441718**
Admire	**0.532467532**	**0.435120435**
Alert	**0.823529412**	**0.554572271**
Cheerful	**0.606060606**	**0.340224454**
Friend	**0.622377622**	**0.656441718**
Peace	**0.78030303**	**0.500806452**
Sad	**0.735294118**	**0.515083799**
client018_a	0.699152542	0.248955224
Client019_g	0.38410596	0.275030902
Client038_b	0.523809524	0.388888889
Client123_b	0.330218069	0.31024735
Client123_l	0.515957447	0.397268777
Client123_m	0.521126761	0.356772334
Client207_b	0.653061224	0.307592472
Client207_c	0.471875	0.267142291

Figure 2. Example of a Dashboard for the association topic/emotion

We are at the moment considering various strategies for visualisation, but some prototypes would for instance show clients(students) grouped by main emotion, and then by topic, so highlighting the main emotion associated to a topic (Figure 2), perhaps flagging clients/students whose emotion is not in line with the general emotion of the cohort. Or,

ISSN 2707-2819

we can explore by individuals, and track their changes over time, to understand whether a change happened, and if this can be associated to a specific topic (Figure 3).

	First Session			Second Session			Third Session			Fourth Session		
	Emotion Compared with GEmotion	Richness Compared with Grichness	POS Compared with GPOS	Emotion Compared with GEmotion	Richness Compared with Grichness	POS Compared with GPOS	Emotion Compared with GEmotion	Richness Compared with Grichness	POS Compared with GPOS	Emotion Compared with GEmotion	Richness Compared with Grichness	POS Compared with GPOS
client004	TRUE	TRUE	FALSE	FALSE	TRUE	FALSE	TRUE	TRUE	FALSE	TRUE	TRUE	TRUE
Client011	FALSE	FALSE	TRUE	FALSE	FALSE	TRUE	FALSE	FALSE	TRUE	FALSE	FALSE	TRUE
Client016	FALSE	TRUE	TRUE	FALSE	TRUE	TRUE	FALSE	TRUE	TRUE	FALSE	TRUE	TRUE
Client019	TRUE	TRUE	TRUE	FALSE	TRUE	TRUE	TRUE	FALSE	TRUE	TRUE	TRUE	TRUE
Client031	FALSE	TRUE	TRUE	FALSE	TRUE	FALSE	FALSE	TRUE	TRUE	FALSE	TRUE	FALSE
Client032	TRUE	FALSE	TRUE	FALSE	FALSE	TRUE	FALSE	FALSE	TRUE	FALSE	FALSE	TRUE
Client034	FALSE	TRUE	TRUE	FALSE	TRUE	TRUE	FALSE	TRUE	TRUE	FALSE	TRUE	TRUE
Client110	FALSE	TRUE	FALSE	FALSE	TRUE	TRUE	TRUE	FALSE	FALSE	FALSE	TRUE	FALSE
Client112	FALSE	FALSE	FALSE	TRUE	TRUE	FALSE	TRUE	TRUE	FALSE	TRUE	TRUE	FALSE
Client123	FALSE	TRUE	FALSE	FALSE	TRUE	FALSE	TRUE	TRUE	TRUE	FALSE	TRUE	FALSE
Client124	FALSE	TRUE	TRUE	TRUE	FALSE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE
Client137	TRUE	TRUE	FALSE	FALSE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE	TRUE
Client219	FALSE	FALSE	TRUE	FALSE	TRUE	FALSE	TRUE	TRUE	FALSE	FALSE	FALSE	TRUE
Client222	TRUE	TRUE	FALSE	FALSE	TRUE	TRUE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE
Client224	TRUE	TRUE	TRUE	FALSE	TRUE	TRUE	TRUE	TRUE	FALSE	FALSE	TRUE	FALSE
Client405	TRUE	TRUE	TRUE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	TRUE
client417	FALSE	TRUE	TRUE	FALSE	TRUE	TRUE	FALSE	FALSE	TRUE	FALSE	FALSE	TRUE
client419	FALSE	TRUE	FALSE	FALSE	TRUE	FALSE	FALSE	TRUE	FALSE	TRUE	TRUE	TRUE
client420	FALSE	TRUE	TRUE	FALSE	TRUE	TRUE	FALSE	TRUE	TRUE	TRUE	TRUE	TRUE
client421	FALSE	TRUE	TRUE	FALSE	TRUE	FALSE	FALSE	TRUE	TRUE	FALSE	TRUE	TRUE

Figure 3. Example of a Dashboard flagging clients who are outliers

Conclusion and future work

The work presented in this paper constitutes the first feasibility study for a system aimed at incorporating an emotional "tracking" mechanism to follow students in a VLE. The feasibility study made use of well-established data analysis, ML techniques and available datasets in order to focus on the selection of tasks that need to be accounted for in the implementation of such a system. Of course, many aspects are still to be developed. First and foremost the level of acceptability/uptake that a system of this sort could have among educators and students. Ethical issues need to be carefully considered before attempting the live deployment of a system of this sort, and we are conducting some focus group experiments with experienced online educators to understand what barriers would prevent the uptake of this solution. From an implementation viewpoint, many aspects need to be investigated further, which might lead to a different decomposition of the tasks. For instance, some automation of the extraction of the topic that caused an emotion arousal could be attempted (Xia & Ding, 2019). And more generally, some dedicated datasets need to be harvested to be able to perform an organic test of the various components. Nevertheless, we believe that this is a worthwhile endeavour which promises to tackle an important health, as well as education, emergency of our times.

References

Abbasi, A., & Chen, H. (2008). Writeprints: A stylometric approach to identity-level identification and similarity detection in cyberspace. *ACM Transactions on Information Systems (TOIS), 26*(2), 7.

Alharbi, L., Grasso, F., & Jimmieson, P. (2019). A Conceptual Framework for Identifying Emotional Factors Leading to Student Disengagement in VLEs. *Proceedings of the 2019 ACM Conference on Innovation and Technology in Computer Science Education*, 293-293.

ISSN 2707-2819

Artino Jr, A. R. (2012). Emotions in online learning environments: Introduction to the special issue. *The Internet and Higher Education, 15.*

Avella, J. T., Kebritchi, M., Nunn, S. G., & Kanai, T. (2016). Learning analytics methods, benefits, and challenges in higher education: A systematic literature review. *Online Learning, 20*(2), 13–29.

Blum, A. L., & Langley, P. (1997). Selection of relevant features and examples in machine learning. *Artificial Intelligence, 97*(1-2), 245–271.

Counselling and Psychotherapy Transcripts, Client Narratives, and Reference Works (2008). Welcome. Alexander Street Publishing. Retrieved from https://asp6new.alexanderstreet.com/psyc/psyc.index.map.aspx

D'Errico, F., Paciello, M., & Cerniglia, L. (2016). When emotions enhance students' engagement in e-learning processes. *Journal of e-Learning and Knowledge Society, 12*(4).

Devasia, T., Vinushree, T., & Hegde, V. (2016). Prediction of students' performance using educational data mining. *Proceedings of the 2016 International Conference on Data Mining and Advanced Computing*, 91–95.

Efklides, A., & Volet, S. (2005). Emotional experiences during learning: Multiple, situated and dynamic. *Learning and Instruction, 15.*

Ekman, P., & Friesen, W. V. (1971). Constants across cultures in the face and emotion. *Journal of Personality and Social Psychology, 17*(2), 124.

Fritz, J. (2011). Classroom walls that talk: Using online course activity data of successful students to raise self-awareness of underperforming peers. *The Internet and Higher Education, 14*(2), 89–97.

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., & Witten, I. H. (2009). The WEKA data mining software: an update. *SIGKDD Explorations, 11*(1), 10-18.

Kantardzic, M. (2011). *Data mining: concepts, models, methods, and algorithms.* John Wiley & Sons.

Krcadinac, U., Pasquier, P., Jovanovic, J., & Devedzic, V. (2013). Synesketch: An open source library for sentence-based emotion recognition. *IEEE Transactions on Affective Computing, 4*(3), 312–325.

Lust, G., Elen, J., & Clarebout, G. (2013). Regulation of tool-use within a blended course: Student differences and performance effects. *Computers & Education, 60*(1), 385–395.

ISSN 2707-2819

Marbouti, F., Diefes-Dux, H. A., & Madhavan, K. (2016). Models for early prediction of at-risk students in a course using standards-based grading. *Computers & Education, 103*, 1–15.

Miller, G. A. (1998). *Wordnet: An electronic lexical database.* MIT press.

Mueen, A., Zafar, B., & Manzoor, U. (2016). Modeling and predicting students' academic performance using data mining techniques. *International Journal of Modern Education and Computer Science, 8*(11), 36.

Pekrun, R. (2006). The control-value theory of achievement emotions: Assumptions, corollaries, and implications for educational research and practice. *Educational Psychology Review, 18*(4), 315–341.

Rollnick, S., & Miller, W. R. (1991). *Motivational Interviewing.* The Guildford Press.

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology, 39*(6), 1161.

Strapparava, C., & Mihalcea, R. (2008). Learning to identify emotions in text. *Proceedings of the 2008 ACM Symposium on Applied Computing*, 1556–1560.

Suh, J. H. (2016). Comparing writing style feature-based classification methods for estimating user reputations in social media. *Springer Plus, 5*(1), 261.

Thorley, C. (2017, September 4). Not By Degrees: Improving student mental health in the UK's Universities. The Progressive Policy Think tank [Blog post]. Institute for Public Policy Research. Retrieved from https://www.ippr.org/publications/not-by-degrees

Whissell, C. (1996). Traditional and emotional stylometric analysis of the songs of Beatles Paul McCartney and John Lennon. *Computers and the Humanities, 30*(3), 257–265.

Xia, R., & Ding, Z. (2019). Emotion-cause pair extraction: A new task to emotion analysis in texts. *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, Florence, Italy*, 1003–1012.

Zheng, R., Li, J., Chen, H., & Huang, Z. (2006). A framework for authorship identification of online messages: Writing-style features and classification techniques. *Journal of the American Society for Information Science and Technology, 57*(3), 378–393

ISSN 2707-2819