

Human context in Sentiment Analysis symbolic technique

Daniel Amo-Filvà¹, Mireia Usart², Carme Grimalt-Álvaro² and Jiahui Chen¹

¹ La Salle, Universitat Ramon Llull, Barcelona, Spain

² Universitat Rovira I Virgili, Tarragona, Spain

Abstract

Learning methodologies in Virtual Learning Environments that encourage students' written communication require additional effort from the trainers, in terms of management and sentimental awareness of both, the group and each participant. Analysing and evaluating sentiment for every message in every conversation is a hard and tedious work. This is one of the reasons why Natural Language Processing (NLP) and Sentiment Analysis (SA) are gaining popularity. The idea of automating the processes of emotional evaluation of students' conversations in an academic context invites us to consider those automatisms as substitutes for manual processes, such as SA. The challenge of including the human context, together with treating the data with adequate privacy in terms of current legislation, makes these techniques complex. There are two main techniques in SA, those based on lexicons and those based on machine learning. In the present study, results of SA based on two different lexicons, are compared with the results of a manual labelling performed by human trainers to test the effectiveness of the SA technique. Regarding the privacy concerns, an open-source local analysis tool was updated and incorporated such automated processes, both for the present study and for trainers to use considering the extracted results. The results show that lexical-based SA processes tend to consider messages towards the extremes (positive/negative), while human beings' evaluation tends towards sentimental neutrality, both in female and male.

Keywords

Sentiment Analysis, Natural Language Processing, Word List, Human Context.

1. Introduction

The use of learning management systems (LMS) has become even more widespread after COVID-19 pandemic [1]. In these virtual contexts, communication between students and trainers is enabled via discussion forums. Conversations regarding learning topics occur asynchronously, and interactions take place naturally considering the cultural, social, and political context of each participant. Emotions always guide interactions [2]. The problem for us is that in face-to-face interactions, we have the non-verbal language that modulates such interactions and provides us with additional information about the meaning and implications of the messages. However, in written communication, it is a more tedious work (you must read each of the communications), and the non-verbal part is missed, which can cause misinterpretation of the messages and teachers/participants not being aware of the state of others. Hence, Sentiment Analysis (SA) can contribute to better assessing participants' emotions and the emotional climate of the classroom to improve teachers' feedback to participants [3, 4]. With emotional climate assessment, trainers can change flow, meaning, and purpose of the participant's conversations to avoid misunderstood and better relations among participants individually and as a group.

The emotional climate of participants is "defined as the quality of the social and emotional interactions between trainees, their peers, and trainers" and is "a key element for creating safe and creative learning environments" [5]. In recent years, trainers have been applying the identification of emotional climate through Big Data techniques considering Machine Learning algorithms or even other Artificial Intelligence approaches running in the cloud. This techno-solutionism raises concerns about

Learning Analytics Summer Institute Spain (LASI Spain) 2022, June 20–21, 2022, Salamanca, Spain

EMAIL: daniel.amo@salle.url.edu (A. 1); mireia.usart@urv.cat (A. 2); carme.grimalt@urv.cat (A. 3); jiahui1@hotmail.es (A. 4)

© 2022 Copyright for this paper by its authors.

Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings (CEUR-WS.org)

ethical practices, privacy and security of data, and context awareness in algorithms. However, analysing textual data in a LMS while teaching can be both stressful and time-consuming for trainers [1], thus, automatic detection, such as SA should be used to help trainers optimize their teaching time.

SA is a natural language processing (NLP) technique that analyses textual data in natural lingua (human readable text) with the aim of extracting and interpreting the expressed emotions. As stated by Usart et al. [5] SA of text input has been characterized as a non-intrusive and behavioural manner of emotion measurement. Hence, a group of participants can be assessed emotionally using this approach to detect the emotional climate and envision results through simple visualizations or dashboards [6–8].

There are two main techniques in SA, those based on lexicons and those based on Machine Learning. In both, as automatic processes, it is very complex to introduce the social, cultural, and political context of human beings. The writing of texts includes different literary figures, such as hyperbole, metaphor, sarcasm, irony, and oxymoron, which semantically, and depending on the context of discourse, can be misinterpreted by automatisms. The challenge of including the human context, together with the challenge of treating the data with adequate privacy in terms of current legislation, makes SA complex and requires advanced approaches such as supervised machine learning. SA based on general lexicon analyses the lexeme of the words in the input texts, which can lead to a distortion in the sentimental result, requiring lexical corpus adapted to the reality to be analysed. The same happens with a machine learning approach, since the training datasets must consider the context analysed. In either case, the real context is complex to incorporate in the automatisms. In summary, SA works using one of two methodologies [9], both facing different challenges such as high-quality data acquisition, human-in-the-loop in machine learning under production mode, human context detection, and ethics and privacy issues:

- Symbolic techniques: The symbolic technique works by breaking a message down into words or sentences and then assigning a sentiment score to each one. A linguistic corpus previously labelled with the sentiments is used to find the sentiment of the analysed texts by applying some sort of combination and aggregation function to the word scoring.
- Machine Learning: The machine learning technique uses a training (fitting) set and a testing set for developing classification models. The training set is previously labeled by humans and used to develop the classification model of sentiments. Results are compared with the test set to validate the model.

In this research, the objective is two-fold. On the one hand, we focus on the development of a tool that allows the execution of SA (symbolic technique). On the other hand, we aim to automatically detect emotions from different learning forums through SA (symbolic technique). Considering that both SA technics face similar problems, we decide to use symbolic technique as a first approach due to three reasons:

1. The development of the tool entails less technological complexity.
2. Scientific literature in education set majority focus on machine learning approaches making the symbolic technique interesting to research.
3. Most of the lexicon-based literature focus on MOOC forums [10]. A manual labelling of a MOOC forum is a tedious and very exhausting work. This could be the reason why no literature was found in educational research considering lexicon-symbolic technique that compares results to manual labelling.

The third reason points out the novelty of our work. Our general aim is to compare SA results with a manual detection made by trainers in a specific educational context. One of the issues that present both SA technics is the difficulty to detect human social and cultural context in natural language. In symbolic techniques, some authors point out the need of an additional domain-focused lexicon to improve analysis accuracy [11, 12]. Comparing results may help to understand how context is considered in both SA techniques in the educational domain considered in this work. May there be important differences between manual and automatic results, which can show that human context is not fully addressed in SA techniques. Also, results may explain possible biases in the labelling of sentiments and thus foster research and development of new domain-focused lexicon. Considering the above, we want to answer how a symbolic technique using a general lexicon database differs from labelling made by humans' experts in pedagogy.

Therefore, it is necessary not only to develop methods that help trainers to process the large amount of textual data generated in LMSs, but also to develop tools that execute those methods, considering

concerns regarding ethics, data privacy, and human context, such as gender, which could help trainers detect differences in participants' interactions caused by gender inequalities. Hence, we will develop tools that execute locally, without the need to send data and process it in the (public) cloud, so any trainer can use it in their own computer. In a next phase, as future work, we will execute the counterpart Machine Learning technique for comparative reasons between automatic and human labelling.

The rest of this paper is organized as follows. Section 2 describes the methodology followed throughout this work. Section 3 presents the results obtained. Section 4 discuss the results. Finally, Section 5 presents the conclusions obtained through this work.

2. Methodology

To achieve the proposed objectives this study is divided into three phases. In the first phase, all datasets are downloaded and prepared for sentiment processing. The datasets are composed of different exported discussion forums from pedagogy courses of Universitat Rovira i Virgili's Moodle platform. In these forums, students discussed topics related to their subject in Spanish. A total of two datasets have been used for the comparison, ranging from 2019 to 2021 courses. The first one, exported from 2019-2020 academic course forum consists of 145 messages and the second, from 2020-2021 forum of 37 messages. A total sample of 182 messages were analysed, both manually and automatically.

In the second phase, jsMLA (JavaScript Moodle Learning Analytics) is adapted and updated [13] This tool can be used for local data analysis in any web browser, which has been developed, scientifically substantiated and published previously. The tool analyses Moodle logs locally with different educational metrics and renders a dashboard with indicator visualizations (see Fig. 1).

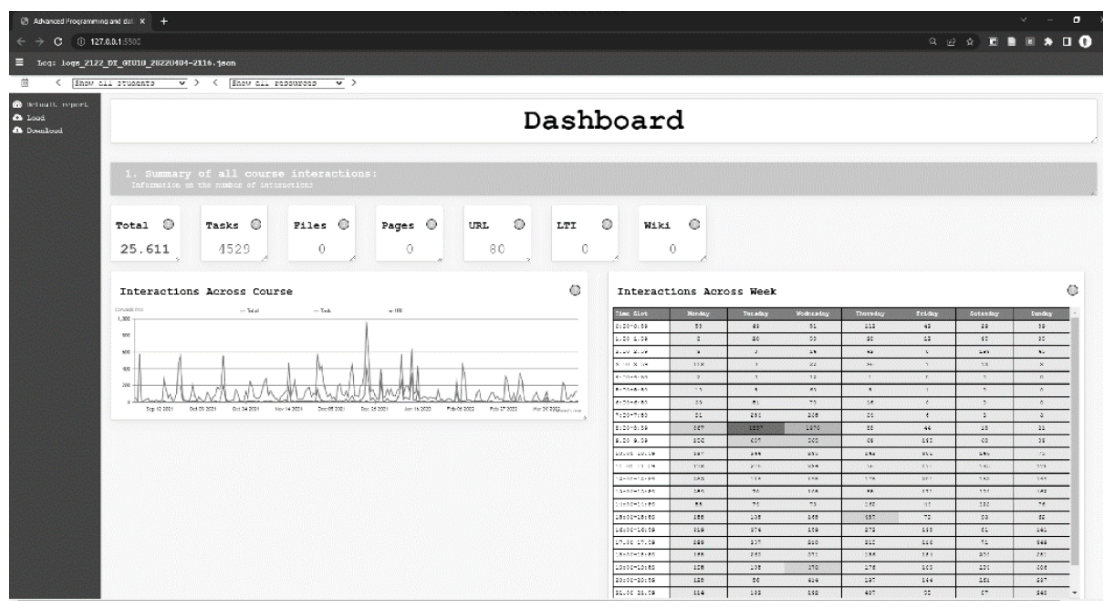


Figure 1: Previous jsMLA dashboard. General information is shown in cards on the top side. Interactions are shown both in graphical and tabular mode.

The new version of the jsMLA tool is renamed to MLA (Moodle Learning Analytics) [14, 15], also developed with web technologies for local deployment to enhance privacy and security in data analysis. The new MLA is served as a standalone multi-platform application to facilitate its distribution, installation with an enhanced user interface for any teacher (view Fig. 2). The newer platform also adds support for Moodle forums logs SA.

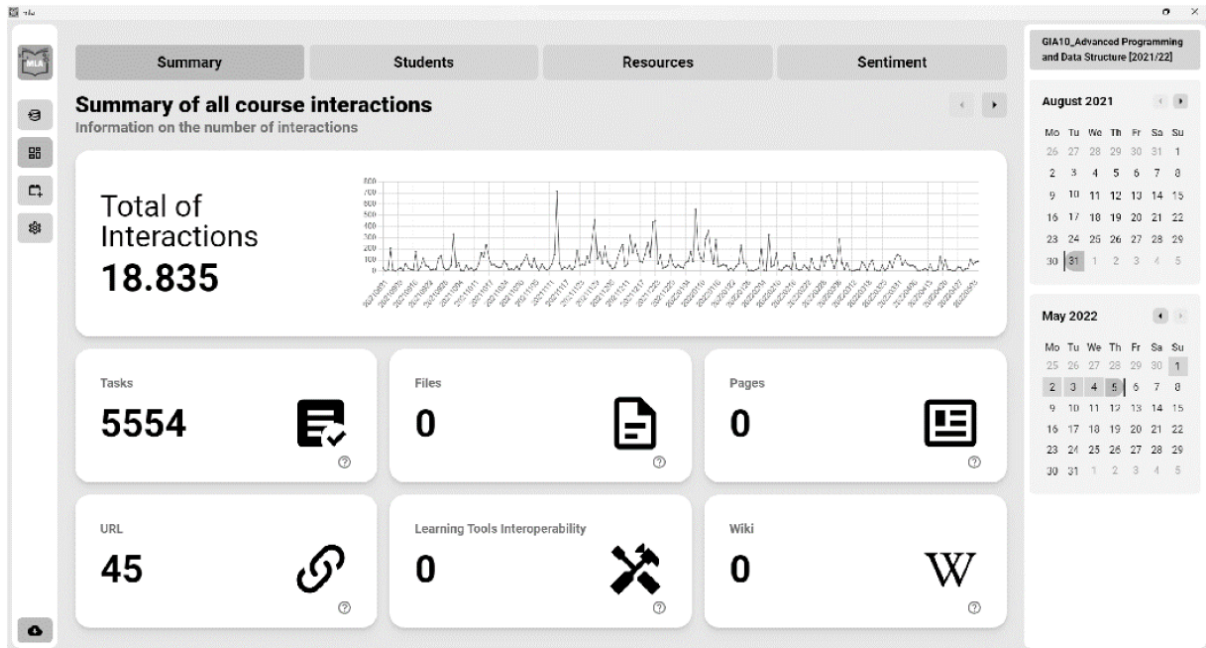


Figure 2: New MLA dashboard.

The process of SA in MLA is divided in different phases (see Fig. 3). Once Moodle forum JSON logs files are imported to the platform, the tool takes the following steps: (i) detects the language of the messages, (ii) performs SA evaluation, (iii) renders the data back to the UI.

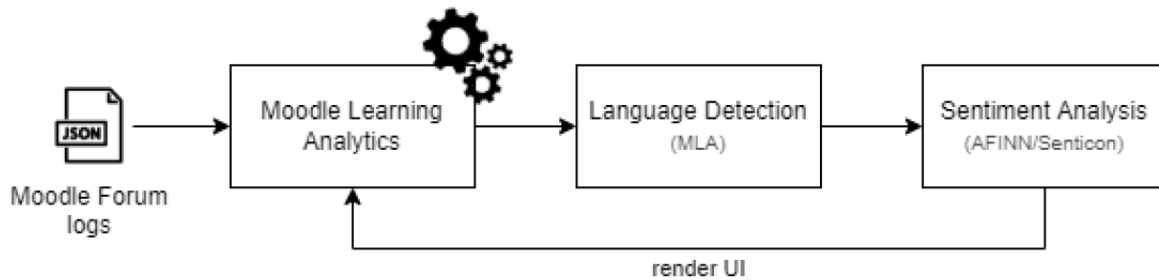


Figure 3: MLA SA processing methodology diagram.

The language used in the forums conditions the possible lexicons, as well as the actual context of the datasets. After exhaustive research, we found no specific lexicon in regards of the context of the datasets. In this respect, the sentiment analysis is implemented with the leverage of AXA's NPL JavaScript [16] library that 1) by default utilizes a lexicon-based approach and 2) allows Spanish words lists. The AXA's NPL library can be configured with 2 different sets of words lists, ML-SentiCon [17] (Multi-Layered, Multilingual Sentiment Lexicon) and a Spanish translated version of the AFINN [18] lexicon. Both lexicon corpus are compared with the human labelling to find out which of the two is more accurate.

In the third phase, both automated and human sentiment analysis is conducted. The forum's messages are classified by gender (M - male / F - female), taking into consideration this perspective in the resulting insights. The labelling comprises the following three types of sentiments, no matter the process of labelling (manual or automatized) (see Fig. 4):

- Positive: messages that transmit an optimistic attitude.
- Negative: messages that are leaned to an adverse thought.
- Neutral: messages that do not fall in either of the previously mentioned categories.

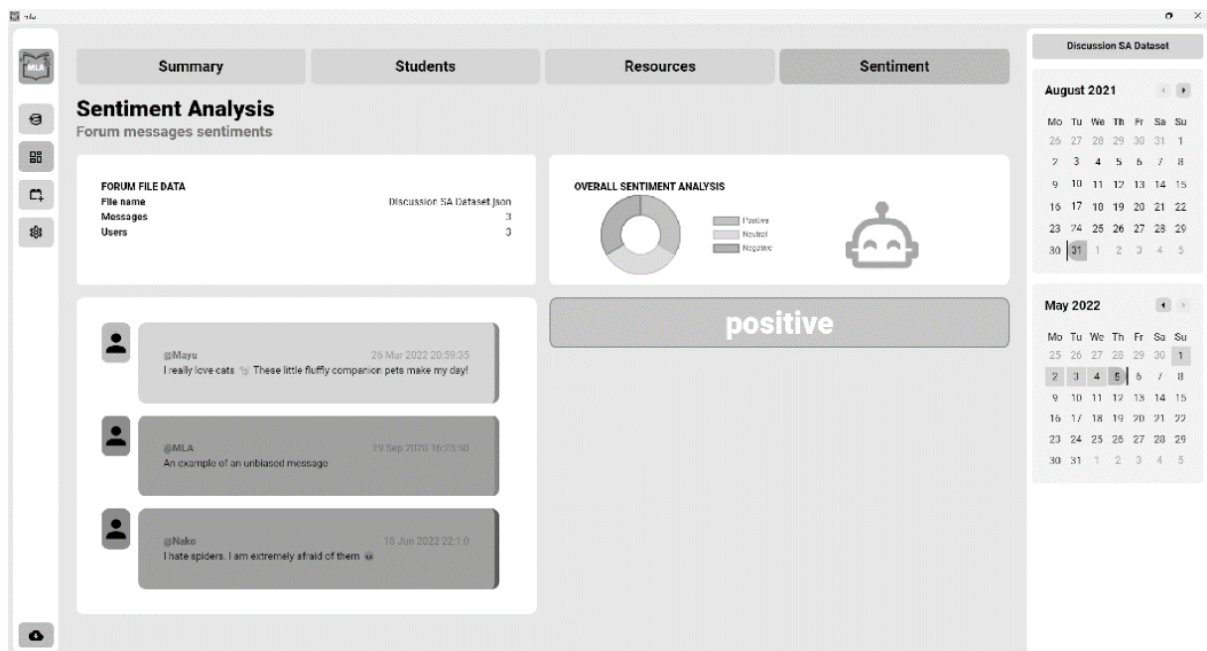


Figure 4: Sentiment Analysis (SA) feature of the MLA tool. Messages are shown in the left bottom section. Positive messages are in green, neutral in yellow and negative in red.

SentiCon uses a valence scoring system that spans from -1 (extremely negative) to +1 (incredibly positive), while AFINN’s ranges, vary from -5 to +5. More positive vocabulary being used in the text means more optimistic attitude and vice versa. Unlike manual classification, humans do this process by considering their contextual background (cultural, social, and political) and the content of the messages themselves. Datasets contain conversations with human-specific nuances that could be overlooked by automatism. As a hypothesis of the study, it is expected that the results will reflect a difference between manual and automated analysis based on lexicons.

3. Results

The results of both manual and automated (using both dictionaries, SentiCon and AFINN) SA are displayed in a vertical bar chart (see Fig. 5). Considering messages processed by human beings:

- The sentiment of the 2019 - 2020 forum is strongly neutral (89.7%), with low positive rating (9.65%) and very low negative rating (0.7%)
- The sentiment of the 2020 - 2021 forum is strongly neutral (75.7%), with a low positive rating (10.8%) and a slight negative rating (13.5%) compared to the previous period

Considering the SentiCon library:

- The sentiment of the 2019 - 2020 forum is strongly positive (68.3%), with substantial negative ratings (31.7%) and no neutral ratings
- The sentiment of the 2020 - 2021 forum is positive (51.4%), with considerable negative rating (29.7%) and neutral (18.9%)

Considering the AFINN library:

- The sentiment of the 2019 - 2020 forum is balanced, positive ratings (49%) and negative (44.8%) with a very low neutral rating (6.2%)
- The sentiment of the 2020 - 2021 forum is slightly more positive (51.4%) than negative (40.5%) with a low neutral rating (8.1%)

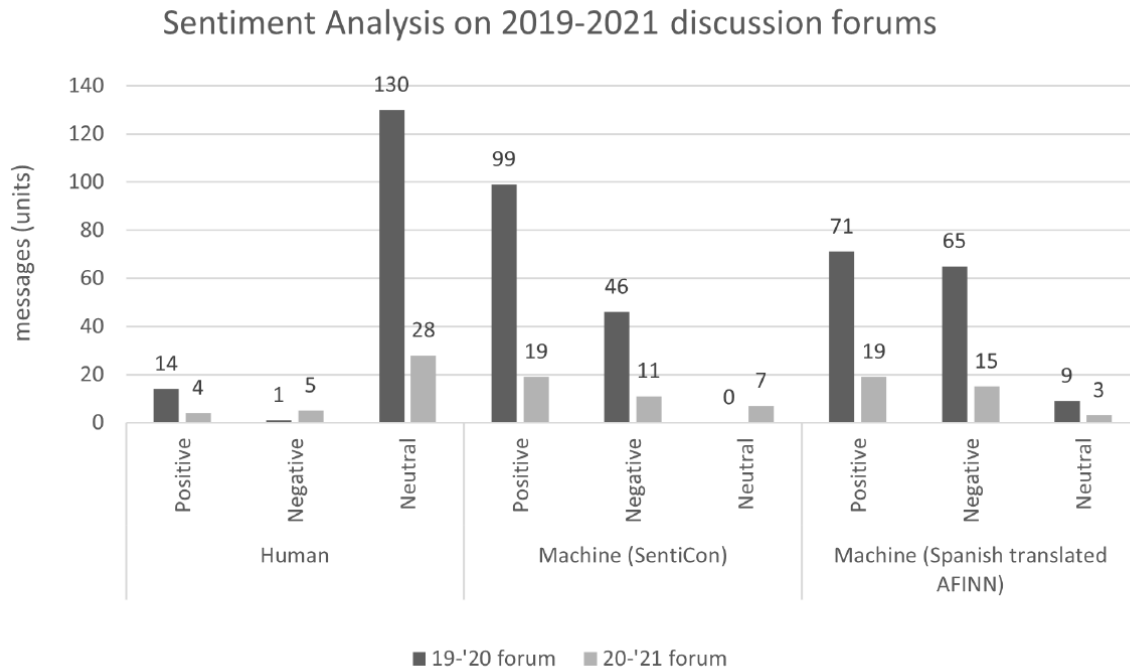


Figure 5: Bar graph of sentiment analysis on '19 - '20 and '20 - '21 discussion forums datasets, a comparison between human vs machine (SentiCon) vs machine (Spanish AFINN).

The Fig. 6 shows the SA results considering the gender of each written message's author. Considering the results for the female gender:

- The sentiment of the 2019 - 2020 forum is strongly neutral (88.5%) in human being ratings, with low positive (10.1%) and exceptionally low negative (1.4%) ratings. The SentiCon based SA tends towards a more positive (69.6%) than negative (30.4%) rating, with no neutrality considered. The AFINN based SA tends towards a low variability ranking between positivity (49.3%) and negativity (43.5%), with an incredibly low neutrality (7.2%).
- The sentiment of the 2020 - 2021 forum is strongly neutral (79.3%) in human being ratings, with low positive (13.8%) and very low negative (6.9%) ratings. The Senti-Con based SA tends towards a more positive (51.7%) than negative (37.9%) rating, with very low neutrality (10.4%). The AFINN based SA tends towards a more positive (51.7%) than negative (31%) rating, with a low neutrality (17.3%).

Considering the results for the male gender:

- The sentiment of the 2019 - 2020 forum is strongly neutral (90.8%) in human being ratings, with few positive ratings (9.2%) and no negative connotations. The SentiCon based SA tends towards a more positive (67.1%) than negative (32.9%) rating, with no neutrality considered. The AFINN based SA tends towards a low variability between positivity (48.7%) and negativity (46%), with an unelevated neutrality (5.3%)
- The sentiment of the 2020 - 2021 forum is more neutral (62.5%) than negative (37.5%) in the human being ratings, with no positive ratings. The SentiCon based SA showed an equally positive (50%) and negative (50%) rating, with no neutral consideration. The AFINN based SA is inclined to a more positive ranking (50%), with equally positive (25%) and neutral (25%) ratings

Sentiment Analysis on female / male messages

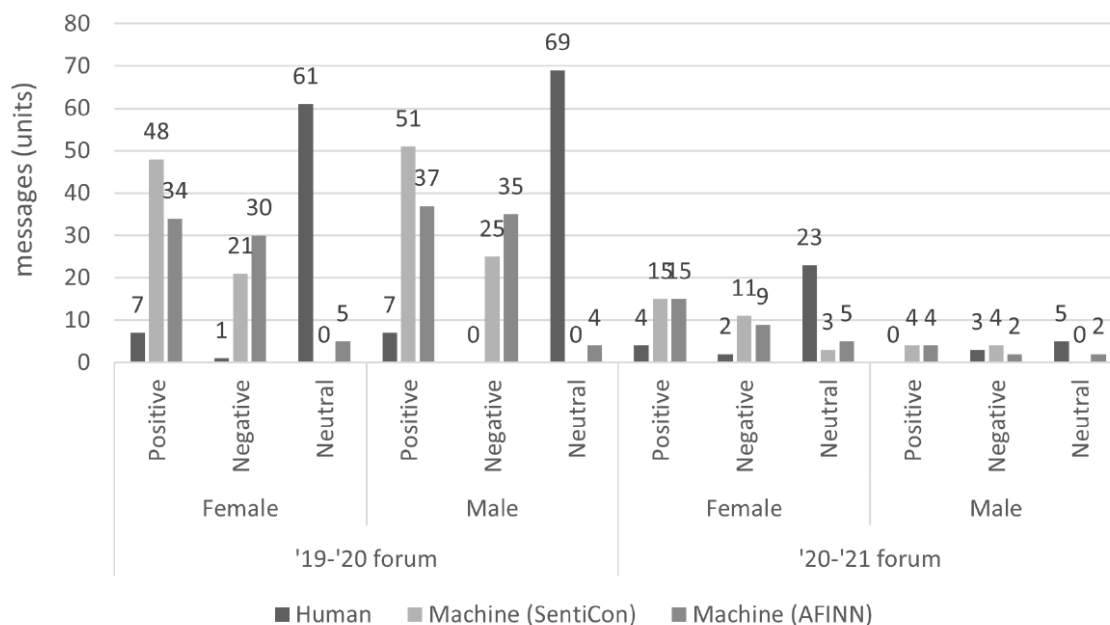


Figure 6: SA bar graph of students' messages performed by human (female vs male), and vs machine (SentiCon) vs machine (Spanish AFINN).

4. Discussion

In light of the results obtained above, human beings are more prone to classify messages as neutral. Changing the lexicon used by the machine to perform NLP, it does not involve significant difference of the labelled sentiment. However, machine's evaluations are more prone to classify messages into either having a cheerful attitude or being negatively assessed. There is a clear difference between humans' sentimental consideration and SA lexicon based.

Another insightful result extracted from this analysis, is the correlation between gender-differentiated messages and the distribution of the diverse types of connotations in these. On the first dataset (2019 – 2020 forum), the distribution of messages sent by females and males are equitable, 69 and 76 messages, respectively. Based on these results, there is no clear evidence whether being female or male may influence in the labelling, though females have (very few) negative considerations as opposed to male.

On the second dataset (2020 – 2021 forum), discussions are fundamentally women predominant with 29 messages, over 8 from males which translates into a ratio of 78% to 22%. Based on these results, there is no clear evidence whether being female or male may influence in the labelling, though male don't show positive consideration compared to women.

Finally, results regarding automatic labelling, indicate that the machine lexicon NLP analysis labelled half of the messages being positive and almost half of the messages being negative. Neutrality is negligible in automatic labelling results.

5. Conclusions

The different LMS-mediated learning methodologies that encourage students' written communication require additional effort on the part of the trainers, in terms of management and sentimental control of both, the group and each participant. Analysing and evaluating sentiment for

every message in every conversation is hard and tedious work. This is one of the reasons why NLP and Sentiment Analysis are gaining increased popularity [5]. The idea of automating the processes of emotional evaluation of certain academic texts invites us to consider them as substitutes for manual processes.

In the present study we compare the results of two automated SA based on two different lexicons (SentiCon [17] and AFINN [18]), with the results of a manual labelling performed by human trainers. For the process to address privacy concerns, we update an opensource local analysis tool [13, 15] and incorporate such automated processes, both for the study and for trainers to use considering the extracted results.

The results of the comparison show that lexical-based SA processes tend to consider messages towards the extremes (positive/negative), with those from the used datasets being considered more positive than negative. The results analysed by human beings tend towards sentimental neutrality, both in those evaluated by female and male, with most messages being marked as neutral and very few as negative/positive. These results can be explained on the one hand by 1) the use of general lexical dictionaries disconnected from the human context in which the conversations are resolved, and 2) the type of activities related to pedagogical studies where the conversations are related to academic work and theory. However, these conclusions do not mean that SA symbolic techniques are biased to the extremes, nor can it be generalized. In particular, these results explain how the lack of context in a dictionary of words can produce bias regarding human expertise of a specific domain context.

In future research we will consider this study's datasets as training datasets to apply machine learning-based SA, to overcome the limitations found regarding the lack of human context in generic lexicon-based SA.

6. Acknowledgments

We appreciate the support received by Silvia Blasi, Aleix Ollé, and Ángel García in the manual labelling process. This project has been funded by the Social Observatory of the “la Caixa” Foundation as part of the project LCF/PR/SR19/52540001.

7. References

- [1] Knopik, T., Osza, U.: E-cooperative problem solving as a strategy for learning mathematics during the COVID-19 pandemic. *Education in the Knowledge Society (EKS)*. 22, e25176 (2021). <https://doi.org/10.14201/eks.25176>
- [2] Bakhtiar, A., Webster, E.A., Hadwin, A.F.: Regulation and socio-emotional interactions in a positive and a negative group climate. *Metacognition Learning*. 13, 57–90 (2018). <https://doi.org/10.1007/s11409-017-9178-x>
- [3] García-Peñalvo, F.J., Corell, A., Abella-García, V., Grande-de-Prado, M.: Online Assessment in Higher Education in the Time of COVID-19. *Education in the Knowledge Society (EKS)*. 21, 26 (2020). <https://doi.org/10.14201/eks.23086>
- [4] García-Peñalvo, F.J., Corell, A., Abella-García, V., Grande-de-Prado, M.: Recommendations for Mandatory Online Assessment in Higher Education During the COVID-19 Pandemic. In: Burgos, D., Tlili, A., and Tabacco, A. (eds.) *Radical Solutions for Education in a Crisis Context: COVID-19 as an Opportunity for Global Learning*. pp. 85–98. Springer, Singapore (2021)
- [5] Usart, M., Grimalt-Álvaro, C., Iglesias-Estradé, A.M.: Gender-sensitive sentiment analysis for estimating the emotional climate in online teacher education. *Learning Environments Research*. 2, 1–20 (2022). <https://doi.org/10.1007/s10984-022-09405-1>
- [6] Álvarez-Arana, A., Villamañe-Gironés, M., Larrañaga-Olagaray, M.: Mejora de los procesos de evaluación mediante analítica visual del aprendizaje. *Education in the Knowledge Society (EKS)*. 21, 13 (2020). <https://doi.org/10.14201/eks.22914>
- [7] Vázquez-Ingelmo, A., García-Peñalvo, F.J., Therón, R.: Towards a Technological Eco-system to Provide Information Dashboards as a Service: A Dynamic Proposal for Supply-ing Dashboards Adapted to Specific Scenarios. *Applied Sciences*. 11, 14 (2021). <https://doi.org/10.3390/app11073249>

- [8] Sarikaya, A., Correll, M., Bartram, L., Tory, M., Fisher, D.: What Do We Talk About When We Talk About Dashboards? *IEEE Transactions on Visualization and Computer Graphics*. 25, 682–692 (2018). <https://doi.org/10.1109/TVCG.2018.2864903>
- [9] Zhang, H., Gan, W., Jiang, B.: Machine Learning and Lexicon Based Methods for Sentiment Classification: A Survey. In: *Proceedings of the 2014 11th Web Information System and Application Conference*. pp. 262–265. IEEE Computer Society, Tianjin, China (2014)
- [10] Mite-Baidal, K., Delgado-Vera, C., Solís-Avilés, E., Espinoza, A.H., Ortiz-Zambrano, J., Varela-Tapia, E.: Sentiment Analysis in Education Domain: A Systematic Literature Re-view. In: Valencia-García, R., Alcaraz-Mármol, G., Del Cioppo-Morstadt, J., Vera-Lucio, N., and Bucaram-Leverone, M. (eds.) *Technologies and Innovation*. pp. 285–297. Springer International Publishing, Cham (2018)
- [11] Yadav, S., Sarkar, M.: Enhancing Sentiment Analysis Using Domain-Specific Lexicon: A Case Study on GST. In: *Proceedings of the 2018 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*. pp. 1109–1114. IEEE, Bangalore, India (2018)
- [12] Muhammad, A., Wiratunga, N., Lothian, R., Glassey, R.: Domain-based Lexicon Enhancement for Sentiment Analysis. In: *Proceedings of the BCS SGAI Workshop on Social Media Analysis 2013 co-located with 33rd Annual International Conference of the British Computer Society's Specialist Group on Artificial Intelligence (BCS SGAI 2013)*. pp. 7–18, Cambridge, UK (2013)
- [13] Amo, D., Cea, S., Jimenez, N.M., Gómez, P., Fonseca, D.: A Privacy-Oriented Local Web Learning Analytics JavaScript Library with a Configurable Schema to Analyze Any Edtech Log: Moodle's Case Study. *Sustainability*. 13, 28 (2021). <https://doi.org/10.3390/su13095085>
- [14] Chen, J., Amo, D.: Moodle Learning Analytics, <https://ls-leda.github.io/Moodle-Learning-Analytics/>
- [15] Amo-Filva, D., Chen, J.: Moodle Learning Analytics. La Salle, URL (2022)
- [16] Seijas, J., Ràfols, R.: NLP.js. AXA (2022)
- [17] Cruz, F., Troyano, J., Pontes, B., Ortega, F.J.: ML-SentiCon: A multilingual, lemma-level sentiment lexicon. *Procesamiento de Lenguaje Natural*. 53, 113–120 (2014)
- [18] Nielsen, F.: AFINN, <http://www2.imm.dtu.dk/pubdb/pubs/6010-full.html>