Artificial Intelligence and Social Computing, Vol. 72, 2023, 151–159
https://doi.org/10.54941/ahfe1003283

Application of Educational Context Data Using Artificial Intelligence Methods

Myriam Peñafiel[1], Maria Vásquez[2], and Diego Vásquez[3]

[1] Escuela Politécnica Nacional, Quito, Ecuador
[2] Universidad Internacional del Ecuador UIDE, Ecuador
[3] Universidad de Las Américas, Quito, Ecuador

ABSTRACT

Today the web generates a large amount of data, the same ones that come from social networks, online platforms, communities, cloud computing, etc., but one type of data has not been recognized for its relevance and that is data from Learning Management Systems like Moodle in the educational context. Considering this context, this research will apply some Artificial Intelligence methods and techniques such as the TSA methodology, Text mining, and Sentiment Analysis to assess the data about the opinion of the students, converting them into stable information structures that allow their reflection and analysis. The work carried out focuses on determining the level of user satisfaction, in this case, the students, of the virtual learning platforms. The results obtained show that applying Artificial Intelligence allows obtaining relevant information that helps to undertake improvement actions by authorities and managers in the educational context based on the opinion of the students, detecting important problems in online learning during these times of COVID-19 we are just past.

Keywords: Artificial intelligence, Education, Opinion mining, Sentiment analysis, Virtual classroom, Students, TSA methodology

INTRODUCTION

The first United Nations World Data Forum made at the beginning of 2017 (Pullinger, 2017) made clear the importance of taking care the most valuable assets in the planet, the data, since there is no benefit in accumulating them, the benefit lies in knowing how to use them. We are at a milestone in the history of humanity it is COVIT-19 pandemic is going through, where our priorities have been totally changed. For this reason, use of data to improve all our processes is essential and one of these is online education for that reason the optimal use of data is critical (Porway, 2020).

This data that needs to be processed since most of it is unstructured data requires advanced techniques to turn it into knowledge, which is the prime, function of Artificial Intelligence (Bag et al., 2022). Thus, data mining arises from the need to obtain knowledge from large volumes of data generated by the development of the web.

The educational context does not escape this need due to the volume of data generated through virtual learning platforms, learning communities, social networks, etc. Thus, educational data mining techniques such as text analysis and sentiment analysis will allow analyzing the data that comes from virtual learning platforms such as Moodle from the students' perspective for data optimization that contributes to the effective decision-making for the improvement of the teaching-learning process, one of the main objectives of the area of educational research (Willaert, 2020).

Therefore, the data that come from the virtual learning environments (Rodrigues et al., 2016) be of special interest for this field to analyze information related to the teaching learning process in institutions of higher education. Especially the data that generated by those involved such as students and teachers to know their opinions, in the same way, which organizations do with their clients (Nassirtoussi, et al., 2014)

Therefore, this work focuses on the application of artificial intelligence to obtain information about the opinions of students and teachers related to their feelings regarding the use of virtual learning environments through methods and techniques of data mining such as text analysis and sentiment analysis.

The sentiment analysis, also called opinion mining, will allow the perception or feeling of the subjective texts of these users to be classified as positive, negative and neutral feelings (Liu, 2012) this research focus in obtain the perception of the use of Virtual classrooms as tools to support the teaching - learning process through the analyze the polarity obtain of the data in the educational context.

In previous works, the authors proposed the evaluation of the use of virtual classrooms by evaluating the closed questions of the surveys (Peñafiel and Lujan, 2014; Peñafiel et al., 2015; Peñafiel et al., 2018a For this, traditional quantitative methods were used, and the need for a method to evaluate open questions was evidenced which due to their subjectivity need other types of evaluation methods.

This paper is an application of data mining using computational techniques such as text mining and sentiment analysis with the objective of evaluate the open questions of the online surveys conducted by students and teachers of the university (Peñafiel et al., 2018b). The results have allowed obtain relevant information in relation to the time that the teachers use when incorporating online platforms in the process of teaching learning, acceptance or rejection of the use of these tools by teachers and students, etc.

This information helps to incorporate improvements in the teaching-learning process, in the use of virtual classrooms, as tools to support face-to-face learning. To validate the proposed, the researchers presented three case studies of interest from the perspective of students and teachers, which are detailed in later sections.

METHOD

Considering that Data Mining allows the preparation of the data so that they can be used in the obtaining of results of the Artificial Intelligence techniques, the use of the TSA Methodology (Peñafiel et al., 2018a; Peñafiel et al., 2018b) that uses the Text Analysis and the Sentiment Analysis for the processing has been selected of the data of this investigation.

Text Mining

The primary objective of text mining is to extract interesting and relevant patterns of behavior to explore knowledge from semi-structured or unstructured textual data (Al-Mahmoud, 2015; Eisenstein, 2019; Chassignol, 2018). The text mining process in schematic form can be said to comprise the following phases: data collection or harvesting, preprocessing, indexing, selection of characteristics, classification, and performance metrics (Hutto and Gilber, 2014). Within text mining tasks (Ferreira-Mello et al., 2019; Greco and Polli, 2020), we can mention some: information retrieval, extraction of concepts, categorization, sentiment analysis, content management, and ontology management.

Sentiment Analysis

The Sentiment Analysis, also called Opinion Mining (Liu, 2012) is an area that has currently attracted the interest of many investigations in recent years, with around 9.000 studies in this field, only in the web of science (Feldman, 2013). Some studies have done literature reviews to determine the most common methods, techniques, and applications in sentiment analysis, just to mention some (Ferreira-Mello et al., 2019; Greco and Polli, 2020). One of the main features is that it allows performing a text analysis of any opinion polling, mainly with a large amount of data in real time. Sentiment Analysis works with the subjective information of human language, so classify those opinions with labels that can be positive or negative so that they can be evaluated.

TSA Metodology

This methodology in Figure 1 will be applied below using data from the opinion of students on the use of virtual classrooms from Learning Management Systems (LMS) like Moodle in the university selected for the case study.

The detail of the applied methodology can be reviewed in previous investigations (Peñafiel et al., 2018a; Peñafiel et al., 2018b), therefore on this occasion it will only be applied following the recommended guidelines.

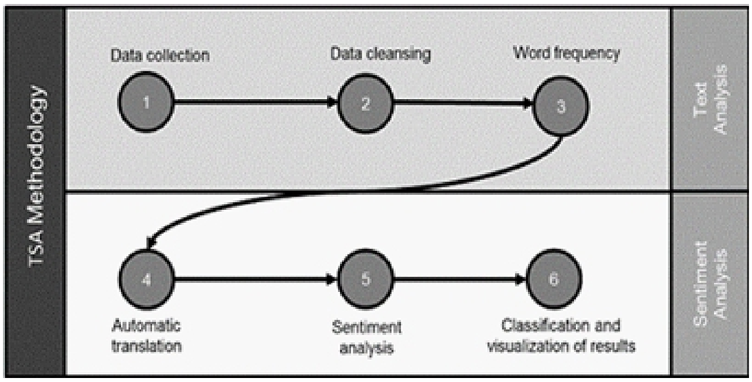

Figure 1: Phases and steps of methodology text analysis sentiment analysis methodology.

RESULTS

At the University of the study case, the students work with virtual classrooms under Moodle and use this tool as a support for face-to-face learning (Willaert, 2020). In this study case, was considered the data of students who used virtual classrooms and answered the survey that was included in the platform in the four semesters between the years 2014 and 2015, the opinion of the teachers you can see at (Peñafiel et al., 2016).

It is worth mentioning that the number of students at the university is 8000 and this survey was sent through the Moodle platform, but since this platform was not mandatory at this time, it was only answered by 610 students. Although by this date thing have changed dramatically due to the COVID-19 pandemic, which we have just experienced, it is important to establish the base point in which this university was, to establish differences, if any, at this date in a future investigation.

In those periods, surveys were applied to students when they started the course (pre-test), and at the end of the course (post-test). Were collected 610 surveys corresponding to the post-test, of which 90,5% of the students answered the open question sent in the survey. This open question was used as input to validate the methodological proposal, in this study case, it was not necessary to select a prominent variable (Miller, 2015) since it was worked on the only open question. The question was:

Question 11: What are the main impediments that you perceived in the access and use of this mode of dictation.

The analysis of open questions is subjective and unquantifiable; this motivated to carry out this research where it is considered a method to evaluate this kind of questions. The results of the first case study of this research aligned to the proposed TSA methodology are presented below:

TSA Phase 1. Text Analysis

In this phase, two experiments were carried out, each with a different technique. In the first experiment, the text analysis was applied without applying lemmatization. For the second experiment, the lemmatization technique was applied, and the lexemes of the most repeated words were considered.

The results obtained with the two techniques can be observed in Table 1. The values obtained in the column of the lemmatization technique will be greater or equal to the values of the normal technique, since this technique groups the whole possible range of words. Also, it should be considered that the lower the frequency, the lower the difference between Lemmatized and Normal.

Next, the results of this table are represented and are observed in Fig. 2, in which it is seen that the count of the Normal text and Lemmatized is similar. The figure shows the ten lemmas most frequent. When comparing the data of the two experiments we can observe that the data present a high correlation

The Pearson correlation was then applied to the two experiments, and it can be concluded that the experiments are statistically equal with a 99.33% correlation The results shows a dispersion of the values obtained

Table 1. Word counting and lemma.

Word/Lemma	Lem	Nor
difficulty/difficulty	700	615
internet/internet	455	453
virtual/virtual	371	371
Problems/problem	282	281
Connect/connect	280	280
computers	177	174
Available/available	174	173
connection/connection	174	174
get/got	173	173
University/univers	173	173
files/archive	170	168
platform	162	162
directories/directori	159	159
understand/understand	85	85
Organized/organized	85	85

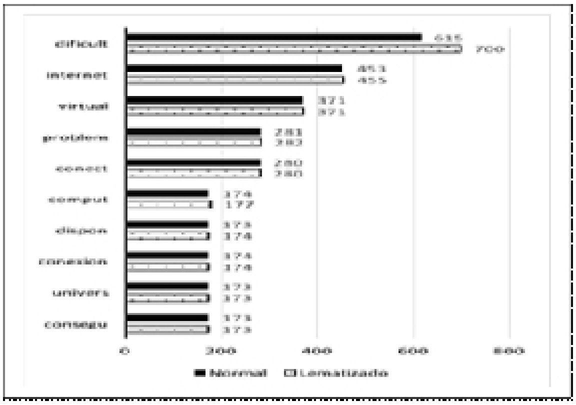

Figure 2: Words frequency.

by the Lemmatized count versus the Normal count. This dispersion diagram represents a linear regression of the two variables.

TSA Phase 2. Sentiment Analysis

To validate the translation process, it was necessary to design two experiments, in experiment 1, where an automatic translation is applied to the data in comparison to experiment 2, where a manual translation is performed, the process performed can be seen in Figure 3.

To do this, once the text was analyzed, we proceeded to look for the most repeated words in the sentences. We considered a representative sample of the 100 responses with the highest frequency, of the total population of 710, which is a valid sample. These words correspond to the 100 words most representative, the same ones must be mapped to their original sentence since

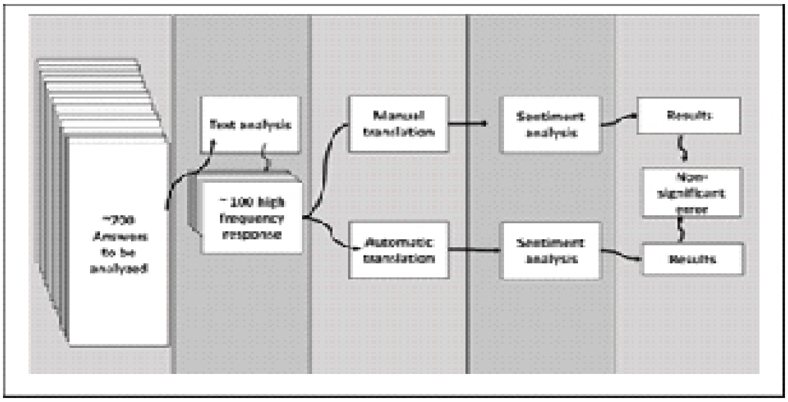

Figure 3: Process diagram for automatic translation validation compared to manual translation.

the original data are used to the process of translation previous to apply the method Sentiment Analysis.

This research carried out for texts in Spanish as well as similar (Brooke and Taboada, 2009) has resorted to automatic translation processes, to validate the proposal was considered the verification of the automatic translation process compared to manual translation estimating that a margin can be introduced of error in the process.

Table 2 shows the results obtained from the analysis of the most representative values used for the experiment 1, which corresponds to the automatic translation, and the experiment 2 corresponding to the manual translation. The distribution of the elements changes, but the percentage of positive, negative and neutral does not change in any of the two experiments, which are maintained with values 11% positive, 17% neutral and 72% negative.

The purpose of these two experiments was to check the accuracy of the translation performed; the results are shown in Table, where the original text, the translated text in the automatic form and the manual translated text of three original phrases taken from the data are presented as examples.

Polarity values of the sentiment analysis of the manual translation in comparison with the automatic translation. As in Phase 1, the Pearson correlation was applied to the two experiments, and we can conclude that the experiments are statistically equal with a 98.07% correlation, therefore the results obtained reflect that the error obtained is not significant.

The results obtained reflect negativity, this is understandable given the nature of the question that allows determining the problems encountered in the use of virtual classrooms as a tool to support the face-to-face learning under the mixed modality.

The most frequent comments detected are listed below:

- "Unorganized virtual classroom, difficulty understanding what has to be done".
- "Internet quality in the classroom".

Table 2. Comparison of manual vs. automatic translation.

Data	Automatic translation	Polarity A. T.	Manual translation	Polarity M. T.
Aula virtual poco organizada, dificultad para entender lo que se tiene que hacer	Unorganized virtual classroom, difficulty to understand what has to be done	−0,076	loosely organized virtual classroom, difficulty understanding what needs to be done	−0,076
Dificultad para conectarse con el aula virtual por problemas con Internet	Difficulty connecting to the virtual classroom due to internet problems	0,0	Difficulty connecting to the virtual classroom problems with Internet	−0,125
Dificultad para conseguir computadoras disponibles con conexión a Internet en la universidad	Difficulty getting computers available with Internet connection at university	0,40	Difficulty getting computers available with Internet connection in college	0,2

- "Difficulty connecting to the virtual classroom due to Internet problems".
- "Difficulty to get available computers with Internet connection at the university".

CONCLUSION

This investigation obtained important results in establishing the base point of this university selected as a case study, as a reference to the state of the use of virtual learning platforms pre-pandemic COVID-19 from the perspective of the students, using methods of Artificial Intelligence.

The Text Mining and Sentiment Analysis (TSA) methodology used Artificial Intelligence techniques to assess the data of the educational context of this case study, as well as it was used in other previous investigations with success, which allowed corroboration its effectiveness.

This research provides results obtained from the student's perspective before COVID-19, where some inconveniences were detected based on the opinion of the students, which reflect problems with didactics and technology.

The information obtained constitutes valuable input for improving the instructional design and detecting the need for guidelines for designing, constructing, and monitoring virtual learning spaces based on good practices and techniques.

This problem leads to the need to train teachers on the subject.

Other problems detected by the students are problems inherent to the technology, considering that the most critical element in the online educational process is the technological infrastructure.

As future work, it is intended to continue experimenting with the TSA methodology to compare these pre-pandemic results with post-pandemic data to assess the impact of this abrupt transition from face-to-face learning to online learning.

The awareness on the part of the academic authorities of the importance of the use of digital media for education as a result of the pandemic we are going through has become evident. Therefore, optimizing the relevant data offered by educational platforms is the responsibility of researchers who can provide relevant information for assertive decision-making to academic authorities, which will benefit all those involved.

REFERENCES

Al-Mahmoud, H. and Al-Razgan M., (2015), "*Arabic Text Mining a Systematic Review" of the Published Literature 2002-2014*. Cloud Computing (ICCC), International Conference pp. 1–7. IEEE.

Bag, S., Srivastava, G., Bashir, M. M. A., Kumari, S., Giannakis, M., and Chowdhury, A. H. (2022). "*Journey of customers in this digital era: Understanding the role of artificial intelligence technologies in user engagement and conversion. Benchmarking:* An International Journal, 29(7), 2074–2098.

Brooke, J. Tofiloski M. and Taboada M., (2009), "*Cross-Linguistic Sentiment Analysis: From English to Spanish"*. In RANLP (pp. 50–54).

Chassignol, M., Khoroshavin, A., Klimova, A., & Bilyatdinova, A. (2018). "*Artificial Intelligence trends in education: a narrative overview"*. Procedia Computer Science, 136, 16–24.

Eisenstein, J. (2019). "*Introduction to natural language processing. Adaptive computation and machine learning series"*. MIT Press, 55, 56.

Feldman, R. (2013), "*Techniques and applications for sentiment analysis. Communications of the ACM"*, 56(4), 82–89.

Ferreira-Mello, R. André, M. Pinheiro, A. Costa E. and Romero, C. (2019), "*Text mining in education*". Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 2019, 9(6), e1332.

Greco, F., and Polli, A. (2020). "*Emotional Text Mining: Customer profiling in brand management"*. International Journal of Information Management, 51, 101934.

Hutto C. J. and Gilbert. E. (2014) "*Vader: A parsimonious rule-based model for sentiment analysis of social media text"*, Eighth international AAAI conference on weblogs and social media.

Liu, B., (2012) "*Sentiment analysis and opinion mining"*. Synthesis lectures on human language technologies, 5(1), 1–167.

Miller, L. D. Soh, L. K. Samal, A. Kupzyk K. and. Nugent, G (2015), "*A comparison of educational statistics and data mining approaches to identify characteristics that impact online learning"*. JEDM-Journal of Educational Data Mining, 7(3), 117–150.

Nassirtoussi, A. K., Aghabozorgi, S., Wah T. Y. and Ngo, D. C. L., (2014)" Text mining for market prediction: A systematic review," Expert Systems with Applications, 41(16), 7653–7670.

Peñafiel, M. and S. Luján-Mora, (2014) "*Uso de aulas virtuales: percepción de los estudiantes en la Escuela Politécnica Nacional de Quito"*, in Actualización de los nuevos sistemas educativos, Estela Bernard Monferrer, Ed. Madrid, España: Asociación Cultural y Científica Iberoamericana, 397–419.

Peñafiel, M., L. Vintimilla, S. Luján-Mora, and P. P Montesdeoca, P. P. (2015) "*Analysis of the usage of virtual classrooms in the National Polytechnic School of Ecuador: Teachers' perception*". Information Technology Based Higher Education and Training (ITHET), International Conference on pp.1–6. IEEE.

Peñafiel, M., Vásquez S. and Luján-Mora S. (2016) "*Use of Virtual Classroom: Summarized Opinion of the Stakeholders in the Learning-Teaching Process*". Proceedings of the 8th International Conference on Computer Supported Education, 1, 314–320, DOI: 10.5220/0005797603140320

Peñafiel, M., Navarrete, R., Lujan-Mora S. and Zaldumbide, J. (2018) "*Bridging the gaps between technology and engineering education*", International Journal of Engineering Education, vol 34 (5), pp. 1479–1494.

Peñafiel, M., Vásquez, S., Vásquez, D., Zaldumbide, J., & Luján-Mora, S. (2018, July). "*Data mining and opinion mining: a tool in educational context*". In Proceedings of the 2018 1st International Conference on Mathematics and Statistics (pp.74–78).

Peñafiel, M., Navarrete, R., Tenemaza, M., Vásquez, M., Vásquez, D., and Luján-Mora, S. (2019). "*The Diffusion of News Applying Sentiment Analysis and Impact on Human Behavior Through Social Media*". In Advances in Artificial Intelligence, Software and Systems Engineering: Proceedings of the AHFE 2019 International Conference on Human Factors in Artificial Intelligence and Social Computing, July 24-28, 2019, Washington DC, USA 10 (pp.250–259). Springer International Publishing.

Porway, J., (2020) UN World Data Forum, Blog, available in: https://unstats.un.org/unsd/undataforum/blog/

Pullinger G., (2017) UN World Data Forum, Bulletin, available in: http://enb.iisd.org/download/pdf/sd/enbplus232num1e.pdf

Rodrigues, R. L Ramos, J. L. C., Silva J. C. S., and Gomes, A. S. (2016), "*Discovery engagement patterns MOOCs through cluster analysis*". IEEE Latin America Transactions, 14(9), 4129–4135.

Willaert, T., Van Eecke, P., Beuls, K., & Steels, L. (2020). "*Building social media observatories for monitoring online opinion dynamics*". Social Media+ Society, 6(2), 2056305119898778.