

Using Sentiment Analysis to Identify Student Emotional State to Avoid Dropout in E-Learning

Míria L. D. R. Bóbó, Federal University of Juiz de Fora, Brazil

Fernanda Campos, Federal University of Juiz de Fora, Brazil

Victor Stroele, Federal University of Juiz de Fora, Brazil*

 <https://orcid.org/0000-0001-6296-8605>

José Maria N. David, Federal University of Juiz de Fora, Brazil

Regina Braga, Federal University of Juiz de Fora, Brazil

 <https://orcid.org/0000-0002-4888-0778>

Tiago Timponi Torrent, Federal University of Juiz de Fora, Brazil

ABSTRACT

Dropping out of school comes from a long-term disengagement process with social and economic consequences. Being able to predict students' behavior earlier can minimize their failures and disengagement. This article presents the SASys architecture based on a lexical approach and a polarized frame network. Its main goal is to define the author's sentiment in texts and increase the assertiveness of detecting the sentence's emotional state by adding author information and preferences. The author's emotional state begins with the phrase extraction from virtual learning environments; then, pre-processing techniques are applied in the text, which is submitted to the complex frame network to identify words with polarity and the author's text sentiment. The flow ends with the identification of the author's emotional state. The proposal was evaluated by a case study, applying the sentiment analysis approach to the student school dropout problem. The results point to the feasibility of the proposal for asserting the student's emotional state and detection of student risks of dropout.

KEYWORDS

E-Learning, Emotional State, Sentiment Analysis, Student Dropout

1. INTRODUCTION

Virtual Learning Environments (VLE) have different models for online interaction, mainly synchronous and asynchronous tools. E-Learning courses allow personal compass and educational resources and services to a large number of students, but they have to comprise a high number of students' dropouts (Queiroga et al., 2017). Discussion forums are among the most common interaction tools offered by VLE, mainly in Massive Online Open Course (MOOCs). The popularity of MOOCs includes

DOI: 10.4018/IJDET.305237

*Corresponding Author

This article published as an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>) which permits unrestricted use, distribution, and production in any medium, provided the author of the original work and original publication source are properly credited.

accessibility to every person who has internet, scalability to handle any number of students with a wide diversity of needs and expectations, and flexibility they provide to learners to study according to their routine. However, students who seek to clarify concepts in online platforms may not get the attention they need, which may favor abandonment (Capuano & Caballé, 2020).

According to Márquez-Vera et al. (2016), dropping out of school comes from a long-term disengagement from school and classes. The student's Motivational Profile is based on detecting characteristics that signal students' demotivation and dropout possibility alerts teachers and tutors to the need for special attention. Therefore, being able to predict this behavior earlier could improve the students' performances, as well as minimize their failures and disengagement (Neves et al., 2021). The students' performances can improve given to the specific opportunities and guidance that they will receive in their trajectory, so educational data is needed to be extracted and analyzed.

Many researchers deal with the combination of intelligent techniques and data extraction from big datasets to develop adaptive e-Learning systems. In this context, Sentiment Analysis (SA) can assist in extracting human thoughts and perceptions from a large amount of data (Hemmatian & Sohrabi, 2019). "It is considered one of the research fields in text mining" (Alatrash et al., 2021). Machine Learning, Lexical, or Hybrid approaches are used in the SA context.

The Sentiment Analysis approach can be influenced by the data type, cost, speed, and method accuracy. According to Hemmatian & Sohrabi (2019), machine learning algorithms are fast and tend to have high accuracy; however, they need human involvement and have high costs due to the large training sets and the algorithms' training time. On the other hand, lexical methods do not require human involvement, are dependent on the lexicon size, and have less accuracy but a low cost of implementation. Because of that, this work focuses on the lexical approach.

One of the biggest challenges of lexical approaches is to deal with semi or unstructured data, which requires Natural Language Processing (NLP) techniques (Hemmatian & Sohrabi, 2019). Furthermore, the dynamism of the language with the increased use of jargon, slang, symbols, images, and emoticons also makes it challenging to identify polarities and the generation of new lexicons. One way to improve the quality of lexical approaches is by using Computational Linguistics solutions, such as FrameNet Brazil (<https://www.ufjf.br/framenetbr/>), which deals with jargon and contexts.

Using only the students' text sentiment is not enough to indicate if they are at risk of school dropout. It is also necessary to analyze the students' Motivational Profile (MP), composed of their engagement, participation, assiduity, dedication, etc. In this way, the student's text sentiment and his Motivational Profile define the student's Emotional State (ES), which can point to the possibility of dropout. Also, knowing the student's Emotional State (is helpful to expand the possibilities of adaptive e-learning (Alatrash et al., 2021; Faria et al., 2017). The solution proposed in this paper is the SASys (Sentiment Analyzes System) architecture, based on a polarized frame network to identify the student's sentiment in texts, with the goal of increasing the assertiveness of detecting his Emotional State. Thus, the main research question is: *Can an architecture, based on a lexical approach, using a polarized frame network, for Sentiment Analysis together with data from the student profile and access to a Virtual Learning Environment, identify the student Emotional State and the risk of school dropout?*

The methodological process followed three main steps:

1. Literature Review to identify the state of the art of Sentiment Analysis. Models, algorithms, techniques, approaches, and flow to be followed were defined.
2. Development step to design and build the architecture to identify early student's dropout possibilities, considering both the student's sentiment in texts and Motivational Profile to identify his Emotional State.
3. Evaluation step to measure the proposed architecture's efficiency in terms of its adherence and accuracy in the education domain.

The article is organized as follows: in section 2, the main concepts are presented, and in section 3, the related work is shown. The architecture and the polarized frame network are presented in section 4. In section 5, the case study and the application of SASys to avoid class dropout are described, and in section 6, the final remarks.

2. BACKGROUND

Sentiment Analysis is the process of extracting human thoughts and perceptions from unstructured texts (Hemmatian & Sohrabi, 2019). It is a multidisciplinary domain that can apply Computational Linguistic, Artificial Intelligence, Text Analysis, or Natural Language Processing techniques to automatically identify the author's sentiment about the text (Altrabsheh et al., 2013; Cirqueira et al., 2018). The student's Emotional State can certainly affect his academic performance, especially if he is unmotivated and disengaged (Liting Xu et al., 2008; Vivian et al., 2016). The student profile must then reflect his Emotional State as much as possible.

According to Altrabsheh et al. (2014), Sentiment Analysis aims to extract the polarity or subjectivity of the text automatically. Polarity can be defined as the judgment that a word brings to mind (Hemmatian & Sohrabi, 2019), and it can be positive, negative, or neutral. Subjectivity represents facts, emotions, visions, or beliefs (Altrabsheh et al., 2013). It is an interdisciplinary area covering solutions ranging from simple Natural Language Processing techniques (segmentation, stemming, tokenization, POST-tagging) to more sophisticated algorithms (Naive Bayes, Support Vector Machine, Maximum Entropy).

The solutions used in Sentiment Analysis can be grouped into three main approaches: Machine Learning, Lexical, and Hybrid (Hemmatian & Sohrabi, 2019). The Lexical approach is based on vocabulary sentiment. Each word has a sentiment or polarity and may or may not have an associated weight (probability of belonging to a class). The polarity is defined by analyzing the word semantic orientation (Machado et al., 2018). Neuro-linguistic programming and Text Analysis techniques are used to identify the sentence syntactic structure and thus define the sentiment of a document or paragraph (Altrabsheh et al., 2014; Hemmatian & Sohrabi, 2019). The basic strategy is to count the number of positive and negative words and associate the polarity with the most significant number (Ding et al., 2008). The use of techniques, such as n-gram, is recurrent in this type of approach.

All approaches can be used in conjunction with the frame network to define the overall sentiment of the text. Given the text to be analyzed, the phrases are forwarded to the polarized frame network, which identifies the evoked frames by the lexical units and returns their polarities. The frame network output can be used as an input parameter for the chosen Sentiment Analysis method.

In lexical approaches, the frame network behaves like the lexicon to be consulted, where words are the lexical units that evoke the frames of the sentence, and the sentiment is the polarity of the frame. The general sentiment can be calculated by the individual sentiments of the lexical units and the valence transformers present in the sentences. Valence transformers are terms that alter the polarity of the closest word that has a sentiment (Taboada et al., 2011). They can be intensifiers (such as a lot, absolutely, too much, repeated exclamation and question marks), reducers (little, less, almost, just) and, negations (no, never).

FrameNet Brasil is a Computational Linguistics laboratory that aims to develop computational solutions to NLP problems using Frame Semantics (Salomão, 2009). Also known as the semantics of comprehension, it gives meaning to words through semantic frames (Fillmore et al., 2003). The project has more than 13,000 lexical units distributed over 1,200 frames. In our research, we used the FrameNet as a lexical solution for the Sentiment Analysis because it has semantics associated with the words in addition to the considerable corpus.

3. RELATED WORK

According to Toti et al. (2021, p. 213), “the application of information extraction and NLP techniques upon unstructured, user-driven texts to derive knowledge, sentiment, and opinion is indeed not novel and has been carried out, per se, both in literature and for commercial applications over the course of at least the last ten years.”. Our approach, besides using NPL, is based on a polarized frame network to identify the student’s sentiment in a text and extracts the student’s profile and access to a VLE to compose the Emotional State.

Dey et al. (2018) proposed Senti-N-Gram, a framework for creating an n-gram lexicon. The authors present the automatic creation of a dictionary of n-gram sentiments through several evaluation databases. The overall sentiment is calculated by having only the difference between positive and negative phrases. Our first approach was based on the 3-gram as Dey et al. (2018), but based on the linguist’s suggestion, the scope has been extended to the nearest punctuation mark, encompassing and influencing all other words on the way, which brought us better accuracy results.

Capuano & Caballé (2020) proposed a multi-attribute text categorization tool that automatically detects useful information from MOOC forum posts, including intents, topics covered, sentiment polarity, level of confusion, and urgency. The extracted information may be used directly by instructors for moderating and planning their interventions as well as input for conversational software agents able to engage learners in guided, constructive discussions through natural language. Our proposal also detects the forum posts’ sentiment (Capuano & Caballé, 2020), but our focus was only on sentiment polarity using the frame network of the FrameNet. The main goal of both works is to engage learners.

Dash et al. (2021) examined the role of communication from users on dropout from digital learning systems. The findings indicate a “main effect of negative sentiment on dropout rates but no effect of positive sentiment on preventing dropout behavior. This main effect is stronger in the early stages of learning and weakens at later stages.”. As our dataset is based on students’ comments in discussion forums, we did not use videos, but different from the authors, we considered the neutral sentiment as well, not only positive and negative ones.

Ullah et al. (2020) proposed an algorithm, a method, and the emoticon lexicon to analyze social media data sentiments. It included machine learning and deep learning algorithms for finding sentiments from twitter-based airline data using several features such as TF-IDF, Bag of words, N-gram, and emoticon lexicons. Our research didn’t include the emoticon, as we first applied to a Virtual Learning Environment. But as future work, we will use emojis and images to detect urgency and confusion (Capuano & Caballé, 2020).

Dr. Murthy et al. (2020) proposed a sentiment classification approach based on Long short-term memory (LSTM) for text data. According to the authors, manual analysis of large amounts of text data is very difficult, so computer processing has emerged. When more training data are available, deep learning methods such as LSTM show better sentiment classification performance with 85% accuracy. In our research, applied to online classes, with texts extracted from the Virtual Learning Environment and social network, the data are less invasive sources and with a more significant emotional source. Our research uses a polarized frame network, which calibration followed a validation process, including teachers and linguist specialists. So, the final assertive brought good results.

Alatrash et al. (2021) presented a method of recommendation model utilizing sentiment analysis based on Convolutional Neural Network (CNN) and Natural Language Processing (NLP) techniques. The research aims towards recommending learning resources relevant to the learners’ preferences with the aid of the previous reviews of other learners, sharing them the top preferences. The authors created their own dataset using Python Script and Scraping of different books’ reviews from Amazon.com. The research of Alatrash et al. (2021) focused on the two-polarity in the classification of sentiment analysis, but, different from the authors, we work with the three-polarity in sentiment analysis classification. We also created our own dataset for training and validation of the architecture. We both adopt a recommender system as the final step to achieve the learner’s efficient learning experience, which is not the focus of this paper.

Considering the works identified in the literature, SA is used to help students in the teaching and learning process and, in some cases, to identify students at risk of dropping out. In addition, some works address other characteristics that, combined with SA, improve the performance of their approaches (Capuano & Caballé, 2020)(Alatrash et al., 2021). However, for the most part, the polarity of the text is the only characteristic considered. Our work considers students' behavioral characteristics in the VLE (interaction, access) and their sentiment in forum posts. We believe that by combining these characteristics, we will be able to identify the Emotional State of the student to reverse possible cases of school dropout.

4. SASYS ARCHITECTURE

SASys architecture detects the author's Emotional State, at a given moment, through the semantic orientation of the words and his interaction in the e-Learning environment. It uses Sentiment Analysis approaches (Lexical, Machine Learning, or Hybrid) to characterize the sentiment.

The flow of detecting the author's Emotional State (indicated by arrows) begins with the data extraction from posts in social networks, forms, or Virtual Learning Environments to detect the sentiment. Next, preprocessing techniques are applied to the texts, such as removing punctuation marks or line breaks. Then the texts are submitted to the complex frame network to identify words with polarity. We describe the network below. After this stage, the text's semantic orientation is identified based on the Sentiment Analysis approaches (Lexical, Machine Learning, or Hybrid). The application defines which approach better fits the solution. The flow ends with identifying the Emotional State of the author of the text, gathering information from the profile and context. (Figure 1).

SASys is a three layers architecture, as we can see in Figure 2:

- **Data extraction:** responsible for identifying the author's data and collecting the most relevant ones for the sentiment analysis and the Emotional State composition. Explicit data are informed by the user and include name, age, and language, and context refers to location, device, etc. The implicit data are not explicitly sent by users, including users' interactions, number of accesses, posts feedback, and text data. The interaction data may include the most accessed web page and the weekday. Feedback expresses the author's sentiment using non-text resources, such as like, love, or follow. The texts are the information used to extract the sentiment that, when added to other collected data, generate the Emotional State of the text's author.
- **Sentiment Analysis:** after collecting the data, it is necessary to use preprocessing techniques before submitting it to a SA solution. The first step is used to clear the initial text of non-relevant data. Then, it aims to model the data according to the algorithm's input, reducing the inconsistency and increasing the method's accuracy (Altrabsheh et al., 2014). The most common preprocessing steps include removing tags, blanks, and stop words, expanding abbreviations, identifying uppercase words, exclamation and question marks, negations, and emoticons, stemming, lemmatization and Term Frequency - Inverse Document Frequency (TF-IDF) (Dey et al., 2017; Haddi et al., 2013; Vijayarani et al., 2015). In this study, we identified uppercase words and used TF-IDF to determine more important words in the text. Also, we removed stop words, as they are not keywords for sentiment analysis.

Any tool for sentiment analysis can be adopted in our architecture proposal, whether it is a Machine Learning, Lexical, or Hybrid approach. According to Hemmatian & Sohrabi (2019), machine learning approaches need voluminous datasets to achieve good accuracy. In this work, we analyze sentiment in Portuguese messages, so the size of the dataset for training was a challenge. Thus, we chose to use a Lexical approach based on the FrameNet.

Figure 1. Workflow of the Emotional State detection

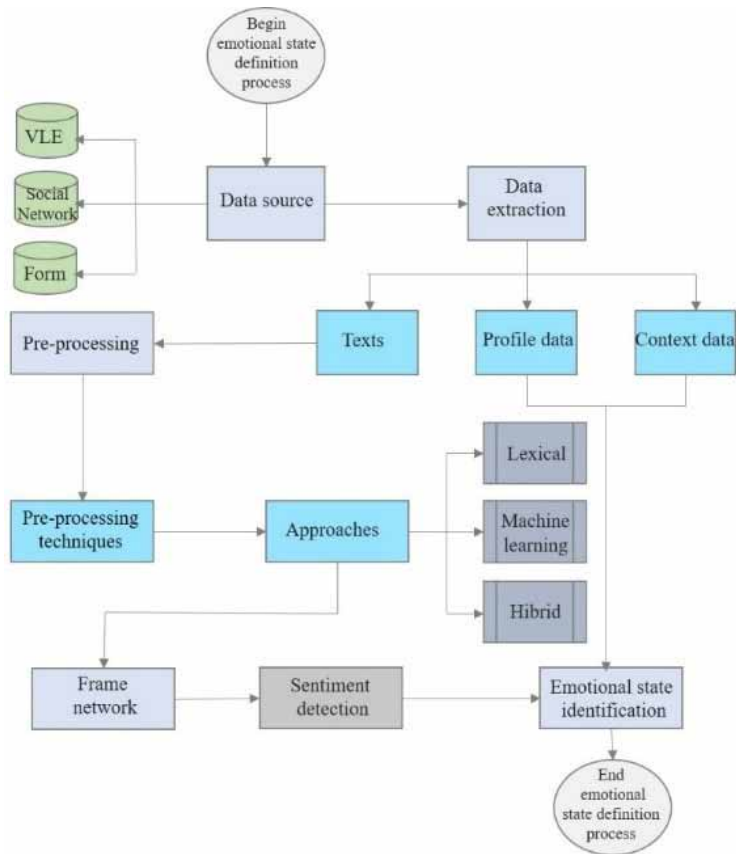
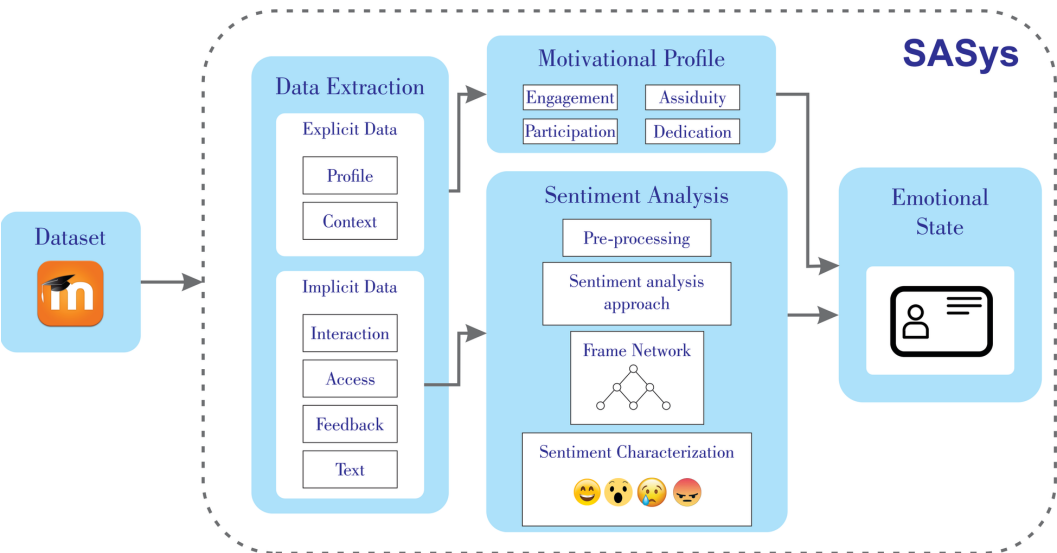


Figure 2. SASys architecture for the virtual learning environment Moodle



After the preprocessing, the polarized frame network identifies the frames evoked by the lexical units and returns their polarities: positive, negative, or neutral (Cirqueira et al., 2018). The frame network output can be used as an input parameter for the chosen Sentiment Analysis approach. The general sentiment can be calculated by considering the individual sentiment of the lexical units (details in Section 4.1) and the valence transformers present in the sentences, based on FrameNet.

In this work, it was used three polarities: positive - words that evoke a positive sentiment, negative - words that evoke a negative sentiment, and neutral - words that do not make a previous judgment. Neutral polarity was used because FrameNet Brasil has many lexical units that evoke scenarios without sentiments (Salomão, 2009).

- **Motivational Profile:** The Motivational Profile is based on extrinsic motivation (Segurado, 2015) by analyzing students' engagement, participation, assiduity, dedication, etc. For example, it is possible to identify students' engagement with the class by analyzing the frequency of access in the Virtual Learning Environment. Explicit and implicit data are submitted to statistical methods to define the student' Motivational Profile based on student engagement (details in Section 5.2).
- **Emotional State identification:** In the proposed architecture, the Emotional State detection happens after the overall sentiment of the text has been generated, together with the profile and context data. The Motivational Profile combined with Sentiment Analysis provides the author's Emotional State when was writing and can be used to recommend personalized information.

SASys architecture was instantiated for the case study in the virtual learning environment Moodle (Figure 2). The implicit data was collected on the students' access and extracted from each forum text. The forums were chosen because they have messages with sentiment content, and they were the communication media among students, teachers, and tutors. Only the students' sentences were considered.

4.1 SASys Polarized Frame Network

An observational study was developed to measure the FrameNet network's accuracy as a lexical approach for SA. The Sentiment Analysis layer was the focus of this phase, leaving the Data Extraction and Emotional State Identification layers to be evaluated later in the case study.

SASys *polarized frame network* is a G graph formed by *frames* and their *lexical units*, and relationships from FrameNet, denoted by the pair $G=[V, E]$, where $V=\{v1, v2,..., vn\}$ is the set of vertices and $E=\{e1, e2,..., em\}$ is the set of arcs. The graph is directed, containing 1,359 vertices and 1,960 edges, and the inheritance relationship allows transitivity and carries semantic information (Salomão, 2009). The transitivity was used in the propagation of the feeling in the network. Figure 3 illustrates the Graph of the FrameNet.

The frames have Lexical Units, which are words with the assigned semantic meaning. We say that the word evokes a frame when the lexical unit assigned to that word is based on the frame. Currently, FrameNet covers more than 13,000 Lexical Units, distributed in more than 1,200 Frames and attested by more than 200,000 annotated sentences.

As the Lexical Units have semantic value, the polarity was attributed to them. To define the polarity, we analyzed the evoked frame's description using the FrameNet web frame management system¹, which allowed the disambiguation of terms by differentiating the use of each of them in different scenarios. For example, the lexical unit leave has a negative polarity when it evokes the abandonment frame (Figure 4).

Once the lexical units' polarity has been defined, we use the *Label Propagation* algorithm (Gregory, 2010) to propagate the polarity to the entire frame network. The algorithm uses topological information to predict the labels of the other nodes in the graph. It assumes that connected nodes should look similar among them and use information from the network structure to determine their similarity.

Figure 3. Graph of the FrameNet network. The size of the vertex was determined by its degree. The colors mark the related components of the graph.

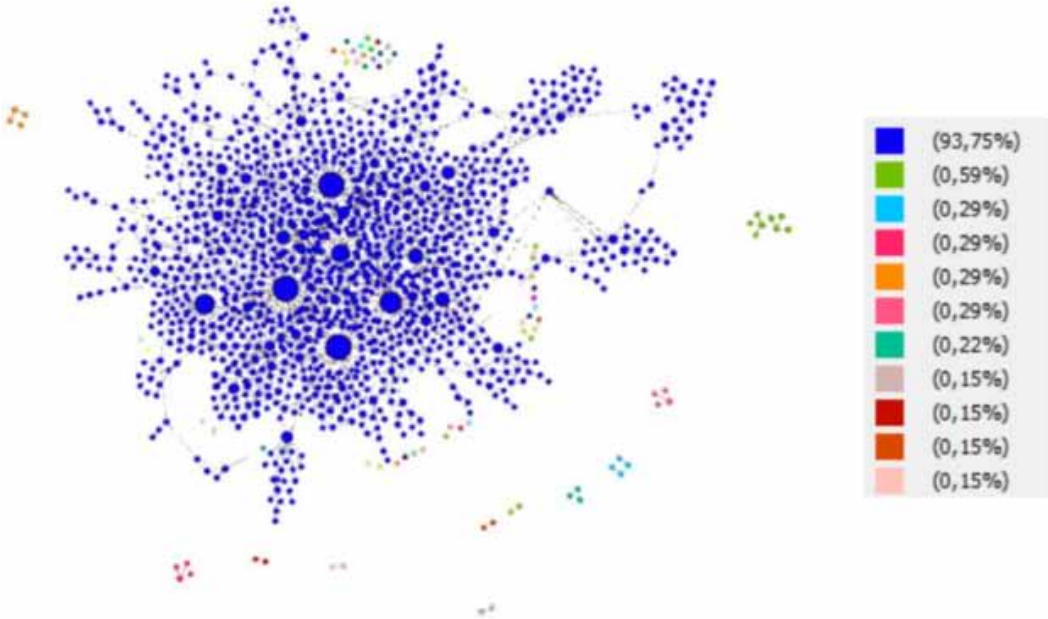
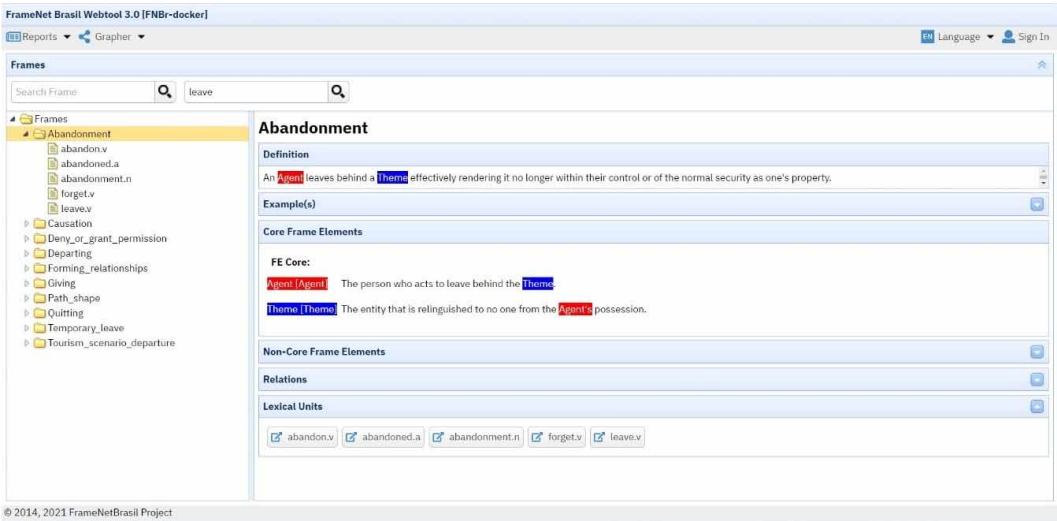


Figure 4. Sample of FrameNet WebTool screen, where the abandonment frame was analyzed



The polarized network performance was assessed using a database² with evaluations of products, films, and restaurants (Kotzias et al., 2015). The database comprises 3,000 labeled phrases taken from three websites: Amazon, IMDb, and YELP. It contains three text files in English, each corresponding to the respective website, with 500 positive sentences and 500 negative ones.

The lexical units of each sentence were identified and passed to the polarized frame network that detected the frames and returned their respective polarities. Thus, given the set of phrases $F=\{f_0,$

..., f_m }, where m is the total of phrases to be analyzed, for all f_i the lexical units were extracted, generating the set $UL(f_i) = \{ul_0, \dots, ul_n\}$, which was passed through the frame network (FN), obtaining the polarity associated with the frame that the lexical unit evoked (n is the number of lexical units of the phrase i). The sentiment of the sentence was defined according to Equation (1), where $FN(ul_j)$ is the polarity of ul_j returned by the frame network defined in Equation (2):

$$P(f_i) = \sum_{j=0}^{|UL(f_i)|} FN(ul_j) \quad (1)$$

$$FN(ul_j) = \begin{cases} 1, & \text{positive} \\ 0, & \text{neutral} \\ -1, & \text{negative} \end{cases} \quad (2)$$

Table 1 presents a sample sentence from the YELP database and the Polarized frame network's sentiment attributed to each of them. The second column shows the lexical units detected in each sentence with the respective polarities of the evoked frames. The last column shows the general sentiment of the sentence.

Table 2 consolidates the results obtained by the proposed approach. The polarized frame network achieved a satisfactory performance for a lexical approach (Hemmatian & Sohrabi, 2019). We use the *unigram* methodology to obtain the sentence sentiment, i.e., we do not use any information of dependence between the words.

The results motivated the use of features that could improve the polarized network performance in real contexts with long sentences, using valence transformers, for example. The next step was to evaluate the SASys architecture in a more comprehensive environment, which could validate all its layers. For this reason, it was submitted to a case study with real context data, which is described in the next section.

Table 1. Sample of sentiment detection from sentences in the YELP database

Sentence	Lexical Unit (polarity)	Sentiment
Our server was very nice, and even though he looked a little overwhelmed with all of our needs, he stayed professional and friendly until the end.	nice (<i>positive</i>), friendly (<i>positive</i>)	Positive
The pizza selections are good.	good (<i>positive</i>)	Positive
I think this restaurant suffers from not trying hard enough.	not (<i>negative</i>), hard (<i>negative</i>), enough (<i>neutral</i>)	Negative

Table 2. Evaluation results for the observational study

Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
86	80	96	87

This step was implemented using the Python language and the NetworkX³ library to create, manipulate, and use complex network functions. After obtaining the polarized network, we stored it in the Neo4j⁴ graph database, allowing graph queries.

5. USING SASYS TO AVOID STUDENTS' DROPOUT

5.1 The Case Study

A case study is suitable for evaluating SASys architecture because it is empirical research carried out in a real context (Dresch et al., 2015). We want to investigate the adequacy of the proposed architecture, the model for detecting the student's Emotional State, and the use of the polarized frame network.

We chose a Distance Education course, as it is a growing application area for Sentiment Analysis (Altrabsheh et al., 2014; Bóbbó et al., 2019; Rani & Kumar, 2017; Shapiro et al., 2017). Some studies show that a student's Emotional State is essential, as it directly impacts the possibility of school dropout (Feng et al., 2016; Lei et al., 2015; Pong-Inwong & Rungworawut, 2014).

From this scenario, the case study's main goal was to identify the students' sentiment from posts in the virtual learning environment, using the polarized frame network, together with information from their profile, and identify the Emotional States with chances of dropping out of school.

5.1.1 Scope

The case study was carried out in a distance education Computer Science Teaching Course from a Federal University in Brazil, whose dataset was used to train and evaluate the proposal. As a sample, we use the information of the Scientific and Educational Research Methodology class in the first semester of 2019. The class had 58 students with the support of the virtual learning environment Moodle. The activities in a group (with no more than three members) were available weekly. This frequency was very important for the data analysis since a single student (in general, the leader) was in charge of posting the activities. Each activity had a specific discussion forum, which posts were used for the Sentiment Analysis. The students' messages were grouped by forum and week. Therefore, we defined seven days for the interaction analysis.

The database for the SA was the texts collected from the forums. The sentiments of the sentences were labeled in two steps. First, they were labeled by two teachers, as usually done by hand, to compare the human expertise in identifying students with negative sentiment with the polarized frame network. In the second step, a linguist evaluated the database to help in the frame network improvement in the automatization process of identifying the student's sentiment. In this case study, the focus was the negative sentiment, which may contribute to school dropout.

The prototype used in this evaluation was developed using an Intel i7-6500U @3.100GHz, 4 CPU Cores, 8GB RAM. Ubuntu 16.04.6 LTS x86 64, Java language, Moodle Web Services⁵, FrameNet API, and PostgreSQL.

Through Moodle Web Services, it was possible to extract access data and all messages posted on the course forums. Data is returned in JSON format. For each student enrolled, the system searches for the messages sent by him in the JSON file and identifies the polarity of the message using the FrameNet API. In addition, student access data is collected considering the file with class access data. This data is stored in PostgreSQL.

After the data of all students are consolidated (access and sentiment), the system performs a statistical analysis on this data to define the thresholds of the Motivational Profiles (details in section 5.2). At this process end, the system defines the Emotional State of each student by considering both the Motivational Profile and the sentiment. Finally, a report is presented to the teacher, showing the student's name, Emotional State, Motivational Profile, and Sentiment Analysis result. Based on this report, the teacher makes the necessary interventions.

5.1.2 Sentiment Analysis Process

The messages were selected from 10 forums as some weeks had no online activities because of face-to-face tests. To facilitate analyzing the messages posted by the students, we defined three sets, each one with 28 messages selected randomly. These sets were used to evaluate the use of FrameNet as a solution for SA. The tables present the average of the results obtained for each set.

After the preprocessing stage, the messages sent by students were selected to be labeled, first, with the expertise of two teachers, and second, with the linguist specialist, which used sentiment detection techniques. This process was necessary to adjust the polarized frame network in terms of assertiveness.

5.1.2.1 Messages Labeled by Distance Education Specialists

Generally, teachers and tutors answer the students' posts in the forums and identify the text's implicit sentiment based on their experience and feelings. With this in mind, our first frame network validation was to repeat this process. The teachers used their expertise to identify characteristics that qualified the phrases within the educational context. They classified the phrases in one of the three polarities: positive, negative, and neutral. Positive phrases should contain praise, inconsistency detection or content mistake (showing cooperation from the student), collaboration among students, content contribution, and all types of motivational subjects pointed out by the student. Negative phrases would be questions about the class content, reports of personal life difficulties that interfere in academic performance, demonstration of frustration, or disorientation. And the neutral ones would be questions about the execution of any task in the Virtual Learning Environment or social messages.

From the messages posted on the forums, 7,4% of them were classified as positive, 48,1% as negative, and 44,5% as neutral. Table 3 shows a sample of each polarity. Among all students, seven were responsible for the generation of all negative content, and three of them had the negative recurrent sentiment, as they published messages with this polarity more than once.

Sentiment Analysis was also performed considering the polarities returned by the frame network with the valence transformers. All the greetings (e.g., good night, good morning, hello) were ignored as preprocessing. The phrases were divided into sentences through the frame network, which returned their polarities.

The comments were divided into sentences, and then their polarities were generated and added to obtain the text's general sentiment. The sum was a binary operation where the following rules determined the general sentiment of the complete comment:

- If the two phrases have equal sentiments, then the recurring sentiment remains.
- If one of the sentences has a negative sentiment, then the negative sentiment remains.
- If one sentence has a neutral sentiment, then the other sentiment remains.

Table 4 illustrates the process of identifying the general sentiment of the text.

Table 3. Sample of phrases labeled by Distance Education specialists

Phrase	Label
The task ends on the 29th, correct? I ask because the task link says that it ends on the 22nd and part of the task says that it ends on the 29 th .	Positive
I'm doing the tasks, but I'm very doubtful if I'm going on the right way, I have a lot of difficulties with this task.	Negative
Good afternoon Teacher, can the task about this week's goal be done in half a page?	Neutral

Table 4. Detection of the comments feeling. Partial polarities are described in numerical format as positive=1, neutral=0, and negative =-1

Phrase	Lexical unit evoked (polarity)	Partial Polarity	Sentiment by the Frame Network
In summary, it is necessary to choose two articles and summarize them, one with quantitative research and another qualitative, citing the sources, or examples of several quantitative and qualitative research models in the field of computing?	necessary (neutral), choose (neutral)	[0,0]	Neutral
The task states that the summary must be on one page. I would like to know if the reference, if necessary, I can put it on a second page. If possible, just to check, could you inform the dimensions used in the standard margins. Thanks.	must(neutral), know(positive), if(reducer not reached), can(positive), if(reducer reached) possible(positive -> negative), could(positive)	[0,+1,0,+1, -1,+1]	Positive
I had the misfortune I could not send the task within the requested time (23:55)... Now it is 23:58h... According to the student's guide orientation, activities with late delivery are evaluated with a 50% discount in the grade; however, the link to send the activity is not available. How can I send the requested task after the deadline?	had(neutral), not(negation reached) could(positive -> negative), not(negation reached), if(reducer reached), available(positive ->negative), can(positive)	[0,-1,-2,+1]	Negative

Table 5 shows the confusion matrix (Shultz et al., 2011) obtained from the True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) measurements. The frame network column refers to the results classification by the polarized frame network.

The confusion matrix was subdivided into three other matrices, one for each polarity. For illustration, we present only the negative matrix (Table 6). Accuracy, precision, recall and, F1-score were calculated separately for each polarity and added up, thus generating the global value for each criterion.

To obtain the frame network's performance values in Sentiment Analysis, the averages of each evaluation criterion were calculated for the labels of the Distance Education specialists (Table 7).

The frame network achieved an accuracy of 63%, with a precision of 46% and a recall of 42% for the Evaluation Criteria of the teachers' label process. According to Hemmatian & Sohrabi (2019), these are reasonable values for a lexical approach. The frame network had better performance at identifying positive polarities, reaching an accuracy of 70%. On the other hand, the precision obtained its greatest result in negative polarity (54%), as it was the polarity with the smallest number of false positives. The recall is associated with the number of samples in the category concerning the base, so the positive polarity had the lowest value.

Table 5. Confusion matrix. Line: label of Distance Education specialists, Column: frame network polarity detection

Label	Frame Network		
	Positive	Negative	Neutral
Positive	1	0	1
Negative	4	7	2
Neutral	3	5	4

Table 6a. Confusion matrix of the negative polarity

Label	Frame Network	
	Negative	Others
Negative	7	6
Others	5	9

Table 6b. Negative polarity evaluation results

Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
59	54	58	56

Table 7. General Evaluation Criteria of the teachers' label process

Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
63	46	42	39

We observed a tendency of the frame network to attribute neutral sentiment to messages. Considering these previous results and improving the frame network's polarity detection process in defining the Sentiment Analysis, the next step was to have the network analyzed by a linguist. The objective was to verify the criteria adopted to define the polarity of the lexical units and, if necessary, to update them.

5.1.2.2 Messages Labeled by a Linguist

The linguist identified new labels for the students' phrases to improve sentiment identification based on the frame network. Table 8 presents a sample of the phrases with the linguist's new labels compared to the teachers' labels.

Most of the sentences labeled as neutral by teachers acquired other polarities when labeled by the linguist. We believe this change is related to the linguistic process that the teachers did not adopt when labeling sentences. The linguist presented suggestions and rules to update the polarity of Lexical Units. Some techniques have been suggested and adopted as rules:

Table 8. Comparative table between the teachers' and the linguist's labels

Phrases	Teachers' label	Linguist's label
In summary, it is necessary to choose two articles and summarize them, one with quantitative research and another qualitative, citing the sources, or examples of several quantitative and qualitative research models in the field of computing?	Neutral	Negative
Teacher, I search for learning objects subject and I found: "According to the Learning Objects Metadata Workgroup, learning objects can be defined by 'any entity, digital or non-digital, that can be used, reused or referenced during learning supported by technologies'. ". Would it be a didactic sequence to teach?	Negative	Negative
I am not good at writing, discussing a text, I did the previous activities, it would be good if they had been evaluated to see how I did, well or awfully bad. Thank you!	Neutral	Positive

1. **Increase in the range of valence transformers:** The previous strategy was based on the 3-gram (Dey et al., 2018), where 3 was the number of words reached by negations, intensifiers, or valence reducers. Its scope has been extended to the nearest punctuation mark, encompassing and influencing all other words on the way.
2. **Change of verb neutrality:** According to the linguist, verbs that express “duty” or “responsibility” have negative sentiment, and those that express “possibility” have a positive sentiment. The change was necessary because the verbs must, need, have, might, choose, among others, were being treated as neutral words and were changed to positive. Another change was related to the words *duty*, *need*, *have to*, which were considered neutral and they should be negative. On the other hand, the words *can*, *choose*, *possibility*, which were also considered neutral words, and they have a positive sentiment.
3. **Change the Citations Marks:** The quotation marks represent the quote and should not be used in sentiment analysis. In the new approach, the number 4 was defined as the minimum number of words for the sentence to be considered a quotation.

A sample of the linguist’s comments and the new classification from the polarized frame network and the updated approach is in Table 9.

Implementing the improvements suggested by the linguist, a new Sentiment Analysis of the students’ comments was obtained using the polarized frame network with the updated rules. The new labeling generated sets of 18,5% positives, 66,7% negatives, and 14,8% neutral phrases. Based on these sets, a new confusion matrix was generated to quantify the new SA approach performance of the frame network. Table 10 shows the results.

The confusion matrix was subdivided into three other matrices, each corresponding to one polarity, and Table 11 illustrates the negative one. Accuracy, precision, recall, and F1-score were calculated separately for each polarity. To obtain the general performance of the new SA approach of the polarized frame network, the averages of each evaluation criterion were calculated.

Table 9. Sample of students’ comments with new labels, defined by the linguist and the polarities detected by the new frame network approach

Comments	Linguist’s label	Frame network label
In the summary it is necessary to choose two articles and make a summary of them, one with quantitative research and another qualitative, citing the sources, or examples of several models of quantitative and qualitative research in the field of computing?	Negative	Neutral
The task ends on the 29th, correct? I ask because the task link says that it ends on the 22nd and part of the task says that it ends on the 29th.	Neutral	Positive
Teacher Fernanda, This activity has to be done right from the Template we are already using ??	Negative	Negative

Table 10. New confusion matrix. Line: Linguist’s label, Column: frame network label.

Label	Frame Network		
	Positive	Negative	Neutral
Positive	5	0	0
Negative	0	14	4
Neutral	1	0	3

Table 11a. Confusion matrix of the negative polarity

Label	Frames Network	
	Negative	Others
Negative	14	4
Others	0	9

Table 11b. Negative polarity evaluation criteria

Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
85	100	78	87

Using the linguist's labeling, the frame network achieved a good overall performance in classifying the students' sentences (Table 12). It obtained its greatest precision in detecting negative phrases because it does not have any false positive in this category, that is, despite not identifying all the negative comments from the base, it did not erroneously categorize phrases from other sentiments as negative. Its greatest results in accuracy, recall, and F1-score were 95%, 100%, and 90%, respectively, in positive sentences, as it achieved greater completeness in this polarity, that is, it was able to recognize all the positive comments present in the database.

The main changes in the polarity of lexical units refer to the range of valence transformers. Table 13 shows samples of these changes showing the Lexical Unit with its polarity. The first is verb neutrality changing, and the others are related to the increase in the range of valence transformers. The valence transformer with the information "reducer not reached" indicates that it has not changed the polarity of the lexical unit that succeeds it.

After updating the Lexical Units polarities, we observed that 57% of the messages changed their polarity from neutral to negative, and 43% remained neutral. On the other hand, 12% of the positive messages changed to neutral, and the rest remained positive. Negative messages have not had their polarity changed.

Table 12. General Evaluation Criteria

Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
88	75	84	78

Table 13. Sample of Lexical Units polarity changing

Teachers process – Lexical Unit(polarity)	Linguist process - Lexical Unit(polarity)
erroneous(negative), can(positive), error(negative), accurate(neutral)	erroneous(negative), can(positive), error(negative), accurate(negative)
separated(neutral), obligatory(negative), if(reducer not achieved), can(positive)	separated(neutral), obligatory(negative), if(reducer achieved), can(positive → negative)
can(positive), better(positive), not(denial not fully achieved), got(positive → negative), understand(positive)	can(positive), better(positive), not(denial achieved), got(positive → negative), understand(positive → negative)

Labeling by teachers carries a sentiment related to the content and progress of the course. In contrast, the linguist's labeling is based on the language fundamentals and does not necessarily take into account the educational context of the messages. The big change in the percentages of neutral labels is related to this difference in the labeling criteria, as several messages did not carry a polarity for the teachers.

5.2 Using SASys to Detect Students at Risk of Dropping Out of the Class

SASys was evaluated to identify students with a chance of dropping out of the class by detecting their Emotional State. Besides the Sentiment Analysis of the text, it was necessary to define the student's Motivational Profile.

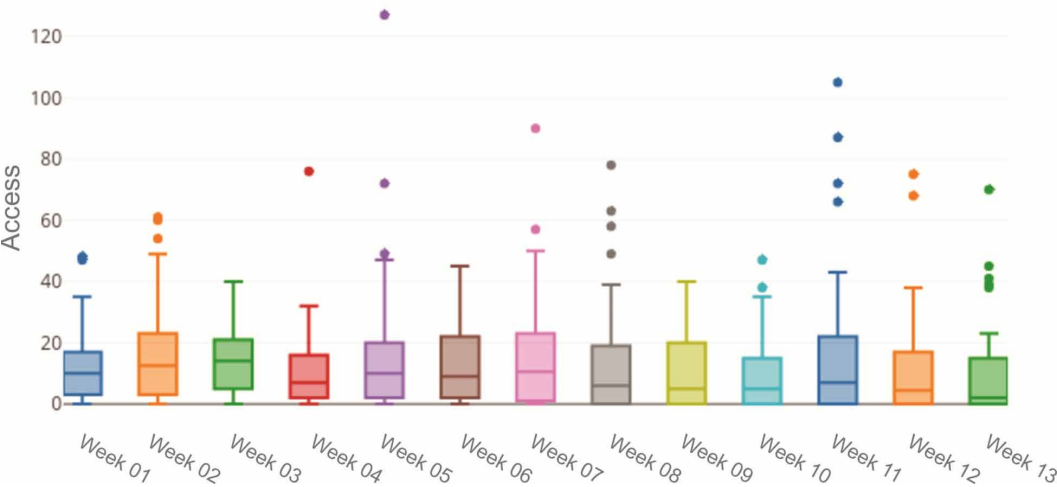
The Motivational Profile was based on students' behavior in the Virtual Learning Environment. It is possible to identify students' engagement with the class by analyzing interaction and access data. The student's Emotional State was defined considering students' engagement to the Virtual Learning Environment and the Sentiment Analysis result. Table 14 shows the data extracted by SASys.

A descriptive data analysis helps understand the students' behavior, considering Interaction and Access data (Table 14). Figure 5 shows the distribution of all students' access per week using boxplots. All graphs have the minimum at zero because at least one student did not access the virtual environment in the corresponding week. The vast majority of outliers represent the groups' leading students. At Week 4, the median value declined, giving evidence of an increase in the number of students with little interaction and, as a result, the existence of unmotivated students with risk of dropout.

Table 14. Moodle data collected by SASys

Profile	Context	Interaction	Access	Feedback	Text
Name	Localization	Accessed pages	Number of accessed pages	Like	Comments
Age	Devices	Weekday	Number of access by day	Follow	Answers
Language	Topics of interest	Visiting hours	Number of accessed topics	Register	Posts
Gender	Preferences	Top topic	Time spent	Share	Questions

Figure 5. Boxplots for students' weekly access in the Virtual Learning Environment



Together with the Sentiment Analysis, these profiles will define the student's Emotional State. Based on the data collected and analyzed in this work, we identified four Motivational Profiles (MP) that stratify students according to the type of access and iteration in the class. The Motivational Profiles defined below include all the students involved in this study. However, we cannot say that these profiles are sufficient for all contexts.

- **MP1 – Enthusiasm:** Students with a large number of actions in the Virtual Learning Environment. These students are more active; they carry out all the proposed class activities and have positive or neutral sentiments. The average access of students in this group was 20 actions per week.
- **MP2 - Moderation and Doubt:** Students with a reasonable number of actions in the VLE. This group comprises two Emotional States: Moderation, students who regularly attend the class, perform most of the proposed activities, and rarely post on forums; and Doubt, students who interact with the forum reporting questions about tasks execution or content. In general, students in this group have neutral and negative sentiments. The average access of students in this group was between 10 and 19 actions per week.
- **MP3 – Warning:** Students with little interaction (dropout alert). These students tend to spend several days (about two weeks) without accessing the Virtual Learning Environment and to perform few activities in the class. Their comments are of disorientation and frustration, indicating the negative sentiment. The average access for this group was less than 10 actions per week.
- **MP4:** Students with almost no interaction (imminent dropout). Students who do not interact with the Virtual Learning Environment and, for a while, their number of actions is null. For this reason, they do not have comments for the Sentiment Analysis. The average access of students in this group was less than 5 actions per week.

Based on the Motivational Profile of the students and the Sentiment Analysis of the comments, three Emotional States of the students were detected in the class dataset:

- **Pleased:** Students MP1 and MP2 who have a positive sentiment.
- **Neutral:** Students MP1 and MP2 who have a negative or neutral sentiment.
- **Unpleased:** PM3 students with negative sentiments.

Table 15 shows the combinations of Motivational Profiles and Sentiment Analysis that generate the students' Emotional States. The students with the Emotional State "Pleased" were engaged and, consequently, pleased with the class. Students with the "Neutral" posted on the discussion forums, clearing up their questions. "Unpleased" students are those who, in addition to having negative comments, have the lowest average access to the Virtual Learning Environment. The Emotional State of some of these students with negative sentiments was withdrawn after their reengagement in the class. They were reclassified as neutral.

Table 15. Motivational Profile and Emotional State

Motivational Profile	Sentiment Analysis of the comments	Emotional State
MP1 e MP2	Positive	Pleased
MP1 e MP2	Neutral	Neutral
MP1 e MP2	Negative	Neutral
MP3	Negative	Unpleased
MP4	-	Unpleased

The students' Emotional State identifies those with a chance to drop out of the online class. Students with MP1 and MP2 and negative sentiment of the comments were considered neutral Emotional State. These students are in a state of avoidance of dropping out because, despite being interested in the class, they have concerns that can lead to demotivation and disengagement. The teacher monitors these students to prevent them from becoming unmotivated, seeking to reverse their negative feeling.

Students with MP3 and MP4 generally have negative sentiments. Therefore, they are identified as students with a chance to drop out, needing special attention from the teacher who intervenes directly with the student to improve his engagement. A motivational limit was established through the median of accesses of each MP capable of identifying unmotivated students. In summary, students with Neutral or Unpleased Emotional States are monitored. The first focuses on changing their sentiment, and the second motivates them to improve their engagement and sentiment.

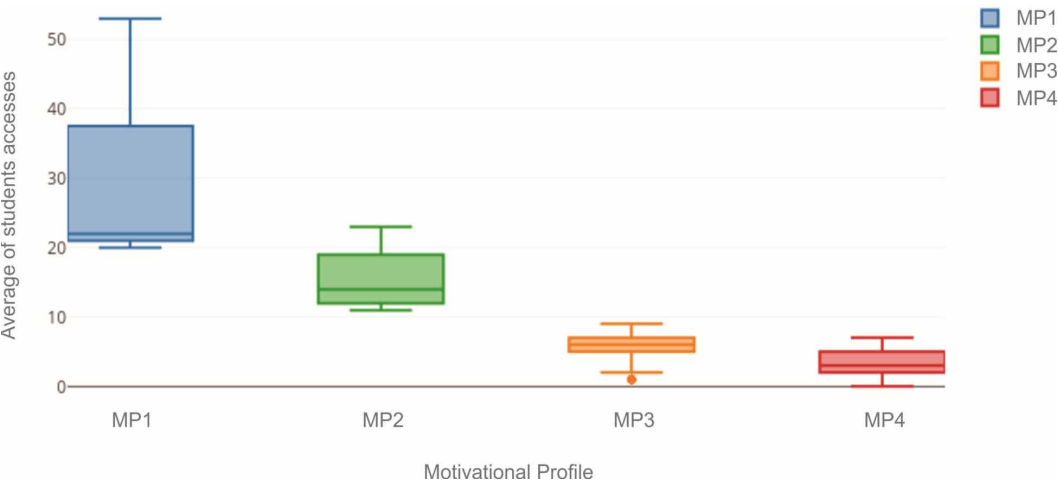
Figure 6 shows a graph where it is possible to observe the distribution of students in their respective MPs. The boxplot medians were used as the motivational limit for each profile: median of MP1=22; MP2=14; MP3=6 and MP4=3. Every student from MP3 and MP4 with an average of access below the median of their respective MP and negative sentiment is also considered unmotivated and at risk of dropout.

When the motivational students' profiles were identified as unmotivated and the Emotional State as unpleased, messages were sent seeking to reverse the tendency to drop out. The students answered an online questionnaire sent in the 11th week through Moodle to detect their preferences, composing the student's context. The questionnaire has seven questions, which sought to discover the most used device for accessing the VLE, preferences regarding the type of content (text, audio, or video), and content size (short or long).

Motivational messages were created for the recommendation. How the messages were presented was defined considering the student's context concerning the type and size of teaching material. Three types of motivational messages were created: video, text, and audio. So, the prototype developed sent messages that adhered to the students' preferences, seeking to reverse the tendency to drop out.

The five students identified as at risk of dropping out received motivational messages. They were sent during the school semester to encourage the students to resume the class activities. It was possible to collect positive feedback because most of the students answered the motivational messages showing interest in recovering the missing activities, except the MP3 student. These students reassessed the Virtual Learning Environment, and they sent the tasks.

Figure 6. Boxplots of the average of students accesses to each motivational profile



In this evaluation, five students were at imminent risk of dropping out due to lack of engagement and negative sentiment about the course. The prototype was able to identify these students, define their Emotional State, alert the teacher about their situation, and send motivational messages. As a result, only one student did not restore his engagement in the course and ended up dropping out. Thus, we can consider that the early identification of demotivated students through SASys and motivational content recommendations can prevent the dropout process.

5.3 Discussion

In e-learning contexts, student dropout is considered one of the main problems and has received much attention from the learning analytics research community, which has reported several approaches to develop models for the early detection of students at risk of dropout (Queiroga et al., 2020). In this case study, it was possible to apply the sentiment analysis in students' texts in an online class and collect their interactions to identify their Motivational Profiles, to create their Emotional State to recognize those at risk of dropout class. The students' texts were from discussion forums, which implied many negative and neutral phrases. The data used in this study allow its expansion and use in other contexts since they are common data in virtual learning environments.

The labeling made by the teachers considered parameters of education such as types of doubts (about the execution of a task or a class content), topics (motivational, social, or collaboration), and expressions (compliments, frustrations, or personal difficulties). This information is based on aspects of the educational context. However, as the SASys architecture is based on a polarized network of frames and not on the identification of the educational context, the results did not accurately identify all sentiments.

Most of the phrases labeled as neutral by the teachers have received other polarities (positive or negative) when labeled by the linguist. With the frame network adjustment, there was then a positive change in the values of the accuracy and precision metrics, showing that the expansion of the range of the reducers improved the performance of the approach.

With the changes suggested by the linguist, the polarized frame network showed a more accurate performance in the Sentiment Analysis of students' texts of an online class, with an accuracy of 88%, as shown in Table 12. According to Larrabee et al. (2019) and Herodotou et al. (2019), one of the main factors involved in accepting learning analytics by teachers and students is the correctness rates involved in the process. As our approach reach a good performance, it can be used during the classes.

Reviewing the main goal of this case study "identify the students' Emotional State from comments in the virtual learning environment, using a polarized frame network, and with information about the Motivation Profile identify students with a chance of class dropout" we can say that:

- Based on the results of the polarized frame network in labeling texts, it can detect the students' sentiments of online classes.
- The data collected through the Virtual Learning Environment were used to define the students' motivational profiles, which together with the SA defined their Emotional States: pleased, neutral, and displeased.
- Based on Emotional State, it was possible to outline some criteria that qualified students as unmotivated or at risk of dropout. These students were notified by the Virtual Learning Environment and received motivational messages.
- There was a good response from the students who received the motivational message, allowing them to return to classes and complete the course.

The results encouraged the use of FrameNet as a dictionary of sentiments in more than one language. However, other evaluations must be conducted to consolidate the architecture proposal, mainly with a larger dataset.

5.3.1 Threats to Validity

We can highlight features of the case study that may have compromised its validity: the areas of the teachers who labeled the phrases in the data set; the amount of data analyzed; the labeling of the phrases, and the evaluation of the Sentiment Analysis approach was done only once; and its performance in numerous sets of texts was not evaluated. In addition, some other threads were identified:

- A sample of lexical units with the sentiment was used as a seed for the label propagation algorithm. This number can impact the polarized frame network's performance, and studies need to be done to define the minimum number of seeds for the algorithm.
- The identification (disambiguation) of the frame evoked by the lexical unit in the sentence to be analyzed is made by the API of FrameNet, so the Sentiment Analysis of the architecture follows the result obtained from this API.
- The frame generation process is still empirical, which may impact the performance of the polarized frame network.
- The definition of Motivational Profiles took into account the data used in this work, and there may be more Motivational Profile types when we consider other contexts.
- The sentiment analysis was not analyzed on how it could help to prevent student's dropout, in depth. Therefore, it is necessary to have a broader dataset.

6. CONCLUSION

This article presented the SASys architecture, instantiated in the context of e-Learning, capable of identifying the students' Emotional State through Sentiment Analysis. The proposal was evaluated through a case study in a real Virtual Learning Environment. The students' motivational levels were identified, the analysis of sentiments was made in the comments of the question forums, and motivational messages were recommended to students with risk of dropout. Students' feedback and interaction gave evidence of the proposal's feasibility in reducing dropout in Distance Education classes. The sentiment analysis using the frame network was first evaluated with a public dataset and proved to be feasible, with an accuracy of 65% and recall of 68%, which are values within the expected results for a lexical approach.

FrameNet was chosen as a SA solution for the proposed architecture, as it has a linguistic basis in the creation and maintenance of frames, which are cognitive structures that define words through scenarios and, for this reason, are also able to identify their sentiment. In the case study, the polarized network of frames achieved accuracy 88%, precision 75%, recall 84%, and F1-score 78%. Such results indicate the use of FrameNet as a lexical solution.

In the case study, SASys was used to define the students' Emotional State, identifying students at risk of dropout. This Emotional State was generated with the data collected in the Virtual Learning Environment together with the SA result, and it was used to recommend motivational messages, showing the proper performance of the architecture.

6.1 Suggestions for Future Work

In this article, we are trying to identify whether sentiment analysis can help detect students at risk of dropping out. In our architecture, sentiment analysis can be performed by any sentiment analysis tool. However, as this work is related to the FrameNet Brazil project, we followed these steps: used the frame network, verified if the network can detect sentiment, and carried out the evaluation to verify if our approach was efficient for detecting students at risk of dropping out, considering the Emotional State detected.

In this way, other evaluations need to be conducted in different conditions and contexts, with other datasets, emojis and images, among others. Future work must evaluate the proposal architecture

from other perspectives, seeking to identify which emotions are associated with students' profiles and how Emotional State changes may suggest changes in students' preferences. In this case, other dataset is necessary to verify more Motivational Profile types. We will also compare the proposed architecture to existing automated sentiment analysis tools to verify its efficiency concerning the others approaches.

ACKNOWLEDGMENT

We would like to thank all the people who participated in the experiments. This study was supported in part by CAPES Foundation (*Coordenação de Aperfeiçoamento de Pessoal de Nível Superior*, Brazil) – Finance Code 001, UFJF, CNPq (435313/2018-5), and FAPEMIG.

CONFLICT OF INTEREST

The authors of this publication declare there is no conflict of interest.

FUNDING AGENCY

The Open Access Processing fee for this article was covered in full by the authors of this article.

REFERENCES

- Alatrash, R., Ezaldeen, H., Misra, R., & Priyadarshini, R. (2021). *Sentiment Analysis Using Deep Learning for Recommendation in E-Learning Domain*. doi:10.1007/978-981-33-4299-6_10
- Altrabsheh, N., Cocea, M., & Fallahkhair, S. (2014). Sentiment Analysis: Towards a Tool for Analysing Real-Time Students Feedback. *Proceedings - International Conference on Tools with Artificial Intelligence*, 419–423. doi:10.1109/ICTAI.2014.70
- Altrabsheh, N., Gaber, M. M., & Cocea, M. (2013). SA-E: Sentiment analysis for education. In *Frontiers in Artificial Intelligence and Applications* (Vol. 255, p. 353–362). doi:10.3233/978-1-61499-264-6-353
- Bóbbó, M., Campos, F., Stroele, V., David, J., & Braga, R. (2019). Identificação do Perfil Emocional do Aluno Através de Análise de Sentimento: Combatendo a Evasão Escolar. *Anais do XXX Simpósio Brasileiro de Informática na Educação (SBIE 2019)*, 1431. 10.5753/cbie.sbie.2019.1431
- Capuano, N., & Caballé, S. (2020). *Multi-attribute Categorization of MOOC Forum Posts and Applications to Conversational Agents*. doi:10.1007/978-3-030-33509-0_47
- Cirqueira, D., Fontes Pinheiro, M., Jacob, A., Lobato, F., & Santana, A. (2018). A Literature Review in Preprocessing for Sentiment Analysis for Brazilian Portuguese Social Media. *Proceedings - 2018 IEEE/WIC/ACM International Conference on Web Intelligence*, 746–749. doi:10.1109/WI.2018.00008
- Dash, R., Ranjan, K. R., & Rossmann, A. (2021). Dropout management in online learning systems. *Behaviour & Information Technology*, 1–15. doi:10.1080/0144929X.2021.1910730
- Dey, A., Jenamani, M., & Thakkar, J. J. (2017). Lexical TF-IDF: An n-gram Feature Space for Cross-Domain Classification of Sentiment Reviews. In *Lecture Notes in Computer Science* (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics): Vol. 10597 LNCS (p. 380–386). Springer. doi:10.1007/978-3-319-69900-4_48
- Dey, A., Jenamani, M., & Thakkar, J. J. (2018). Senti-N-Gram: An n-gram lexicon for sentiment analysis. *Expert Systems with Applications*, 103, 92–105. doi:10.1016/j.eswa.2018.03.004
- Ding, X., Liu, B., & Yu, P. S. (2008). A holistic lexicon-based approach to opinion mining. *WSDM'08 - Proceedings of the 2008 International Conference on Web Search and Data Mining*, 231–239. doi:10.1145/1341531.1341561
- Dresch, A., Lacerda, D. P., & Antunes Jr, J. A. V. (2015). *Design science research: A method for science and technology advancement*. 10.1007/978-3-319-07374-3
- Faria, A. R., Almeida, A., Martins, C., Gonçalves, R., Martins, J., & Branco, F. (2017). A global perspective on an emotional learning model proposal. *Telematics and Informatics*, 34(6), 824–837. doi:10.1016/j.tele.2016.08.007
- Feng, Y., Chen, D., Chen, H., Wan, C., & Xi, P. (2016). The assessment of the points reward mechanism in online course forum. *PIC 2016 - Proceedings of the 2016 IEEE International Conference on Progress in Informatics and Computing*, 722–727. doi:10.1109/PIC.2016.7949593
- Fillmore, C. J., Johnson, C. R., & Petruck, M. R. L. (2003). Background to framenet. *International Journal of Lexicography*, 16(3), 235–250. doi:10.1093/ijl/16.3.235
- Gregory, S. (2010). Finding overlapping communities in networks by label propagation. *New Journal of Physics*, 12(10), 103018. doi:10.1088/1367-2630/12/10/103018
- Haddi, E., Liu, X., & Shi, Y. (2013). The role of text preprocessing in sentiment analysis. *Procedia Computer Science*, 17, 26–32. doi:10.1016/j.procs.2013.05.005
- Hemmatian, F., & Sohrabi, M. K. (2019). A survey on classification techniques for opinion mining and sentiment analysis. *Artificial Intelligence Review*, 52(3), 1495–1545. doi:10.1007/s10462-017-9599-6
- Herodotou, C., Rienties, B., Verdin, B., & Boroowa, A. (2019). Predictive learning analytics ‘at scale’: Towardsguidelines to successful implementation in Higher Education based on the case of the Open University UK. *Journal of Learning Analytics*. Advance online publication. doi:10.18608/jla.2019.61.5
- Kotzias, D., Denil, M., De Freitas, N., & Smyth, P. (2015). From group to individual labels using deep features. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 597–606. doi:10.1145/2783258.2783380

- Larrabee Sønderlund, A., Hughes, E., & Smith, J. (2019). The efficacy of learning analytics interventions in higher education: A systematic review. *British Journal of Educational Technology*, 50(5), 2594–2618. doi:10.1111/bjet.12720
- Lei, C. U., Hou, X., Kwok, T. T. O., Chan, T. S. F., Lee, J., Oh, E., Gonda, D., Yeung, Y. C. A., & Lai, C. (2015). Advancing MOOC and SPOC development via a learner decision journey analytic framework. *Proceedings of 2015 IEEE International Conference on Teaching, Assessment and Learning for Engineering, TALE 2015*, 149–156. doi:10.1109/TALE.2015.7386034
- Machado, M. T., Pardo, T. A. S., & Ruiz, E. E. S. (2018). Creating a Portuguese Context Sensitive Lexicon for Sentiment Analysis. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*: Vol. 11122 LNAI (p. 335–344). doi:10.1007/978-3-319-99722-3_34
- Márquez-Vera, C., Cano, A., Romero, C., Noaman, A. Y. M., Mousa Fardoun, H., & Ventura, S. (2016). Early dropout prediction using data mining: A case study with high school students. *Expert Systems: International Journal of Knowledge Engineering and Neural Networks*, 33(1), 107–124. doi:10.1111/exsy.12135
- Murthy, Allu, Andhavarapu, & Bagadi. (2020). Text based Sentiment Analysis using LSTM. *International Journal of Engineering Research and*, 9(5). 10.17577/IJERTV9IS050290
- Neves, F., Campos, F., Ströele, V., Dantas, M., David, J. M. N., & Braga, R. (2021). Assisted education: Using predictive model to avoid school dropout in e-learning systems. In *Intelligent Systems and Learning Data Analytics in Online Education* (pp. 153–178). Elsevier. doi:10.1016/B978-0-12-823410-5.00002-4
- Pong-Inwong, C., & Rungworawut, W. S. (2014). Teaching senti-lexicon for automated sentiment polarity definition in teaching evaluation. *Proceedings - 2014 10th International Conference on Semantics, Knowledge and Grids, SKG 2014*, 84–91. doi:10.1109/SKG.2014.25
- Queiroga, E., Cechinel, C., & Araújo, R. (2017). Predição de estudantes com risco de evasão em cursos técnicos a distância. *Anais do XXVIII Simpósio Brasileiro de Informática na Educação (SBIE 2017)*, 1, 1547. 10.5753/cbie.sbie.2017.1547
- Queiroga, E. M., Lopes, J. L., Kappel, K., Aguiar, M., Araújo, R. M., Munoz, R., Villarroel, R., & Cechinel, C. (2020). A Learning Analytics Approach to Identify Students at Risk of Dropout: A Case Study with a Technical Distance Education Course. *Applied Sciences (Basel, Switzerland)*, 10(11), 3998. doi:10.3390/app10113998
- Rani, S., & Kumar, P. (2017). A Sentiment Analysis System to Improve Teaching and Learning. *Computer*, 50(5), 36–43. doi:10.1109/MC.2017.133
- Salomão, M. M. M. (2009). FrameNet Brasil: Um trabalho em progresso. *Calidoscópio*, 7(3), 171–182. doi:10.4013/cld.2009.73.01
- Segurado, R. M. (2015). *Gestão de recompensas e perfil motivacional em profissionais de TI: papel na prevenção da rotatividade externa - Iscte – Instituto Universitário de Lisboa*. <https://www.iscte-iul.pt/tese/606>
- Shapiro, H. B., Lee, C. H., Wyman Roth, N. E., Li, K., Çetinkaya-Rundel, M., & Canelas, D. A. (2017). Understanding the massive open online course (MOOC) student experience: An examination of attitudes, motivations, and barriers. *Computers & Education*, 110, 35–50. doi:10.1016/j.compedu.2017.03.003
- Shultz, T. R., Fahlman, S. E., Craw, S., Andritsos, P., Tsaparas, P., Silva, R., Drummond, C., Ling, C. X., Sheng, V. S., Drummond, C., Lanzi, P. L., Gama, J., Wiegand, R. P., Sen, P., Namata, G., Bilgic, M., Getoor, L., He, J., Jain, S., & Mueen, A. et al. (2011). Confusion Matrix. In *Encyclopedia of Machine Learning* (pp. 209–209). Springer US. doi:10.1007/978-0-387-30164-8_157
- Taboada, M., Brooke, J., Tofiloski, M., Voll, K., & Stede, M. (2011). Lexicon-based methods for sentiment analysis. *Computational Linguistics*, 37(2), 267–307. doi:10.1162/COLI_a_00049
- Toti, D., Capuano, N., Campos, F., Dantas, M., Neves, F., & Caballé, S. (2021). Detection of Student Engagement in e-Learning Systems Based on Semantic Analysis and Machine Learning. In *Lecture Notes in Networks and Systems* (Vol. 158, pp. 211–223). LNNS. doi:10.1007/978-3-030-61105-7_21
- Ullah, M. A., Marium, S. M., Begum, S. A., & Dipa, N. S. (2020). An algorithm and method for sentiment analysis using the text and emoticon. *ICT Express*, 6(4), 357–360. doi:10.1016/j.icte.2020.07.003

Vijayarani, S., Ilamathi, M. J., & Nithya, M. (2015). Preprocessing Techniques for Text Mining - An Overview. *International Journal of Computer Science & Communication Networks*, 5(1), 7–16.

Vivian, R., Falkner, K., Falkner, N., & Tarmazdi, H. (2016). A method to analyze computer science students' teamwork in online collaborative learning environments. *ACM Transactions on Computing Education*, 16(2), 1–28. doi:10.1145/2793507

Xu, L., Zhan, Y., & Zhu, M. (2008). A guide based on situation and emotion analysis in ontology-based E-Learning system. *2008 3rd IEEE Conference on Industrial Electronics and Applications*, 2417–2421. doi:10.1109/ICIEA.2008.4582950

ENDNOTES

- ¹ <http://webtool.framenetbr.ufjf.br/>
- ² <https://www.kaggle.com/marklvi/sentiment-labelled-sentences-data-set/data>
- ³ <https://networkx.github.io/>
- ⁴ <https://neo4j.com/>
- ⁵ https://docs.moodle.org/dev/Web_service_API_functions