Original Article

ISSN 2177-8310
DOI: 10.18264/eadf.v12i1.1732

Application of an Emotion Miner in Discussion Forums of a Massive Open Online Course (MOOC) in Brazil: An Approach Using the Naive Bayes Algorithm

Aplicação de um Minerador de Emoções em Fóruns de Discussão de um Massive Open Online Course (MOOC) Brasileiro: Uma Abordagem Utilizando o Algoritmo Naive Bayes

Vanessa Faria de Souza*

Instituto Federal de Educação, Ciência e Tecnologia do Rio Grande do Sul - Campus Ibirubá – Ibirubá – RS – Brasil.

*vanessa.souza@ibiruba.ifrs.edu.br

Abstract

MOOCs are gradually evolving which is due to the wide dissemination of virtual learning environments, which provide means of interaction for participants, one of which is the discussion forum, which has a lot of information about student engagement. However, reading all the posts is a difficult task, as MOOCs tend to have a very high number of students enrolled. In this sense, text mining can help teachers gain relevant knowledge about students' posts. Thus, in this study, an emotion miner was implemented for MOOC forums, using the Python programming language, in order to identify and analyze the feelings that each student expresses when interacting with others, in these environments. The results obtained, in initial experiments, show that the miner proved to be efficient in extracting students' emotions, reaching an accuracy of 40% and that positive feelings such as joy and surprise reflect on the conclusion of the MOOCs, while negative feelings such as sadness and anger are indicative of dropping out of the course.

Keywords: MOOCs. Discussion forums. Miner of emotions. Naive Bayes Algorithm.

Received: 01/25/2022
Accepted: 06/06/2022
Published: 06/07/2022

COMO CITAR ESTE ARTIGO

ABNT: SOUZA, V. F. Aplicação de um Minerador de Emoções em Fóruns de Discussão de um Massive Open Online Course (MOOC) Brasileiro: uma Abordagem Utilizando o Algoritmo Naive Bayes. **EaD em Foco**, v. 12, n. 2, e1732, 2022. doi: https://doi.org/10.18264/eadf.v12i2.1732

Aplicação de um Minerador de Emoções em Fóruns de Discussão de um Massive Open Online Course (MOOC) Brasileiro: Uma Abordagem Utilizando o Algoritmo Naive Bayes

Resumo

Em gradativa evolução devido à disseminação dos ambientes virtuais de aprendizagem, os MOOCs disponibilizam aos participantes inúmeros meios de interação. Dentre esses meios, destaca-se o fórum de discussão, ambiente que registra diferentes informações a respeito do engajamento dos alunos. Contudo, realizar a leitura de todas as postagens é uma tarefa difícil, pois os MOOCs costumam ter uma faixa muito alta de alunos matriculados. Nesse sentido, a mineração de textos pode auxiliar professores a obter conhecimentos relevantes sobre as postagens dos alunos. Levando em consideração essas discussões, neste estudo, foi realizada a implementação de um minerador de emoções para fóruns MOOC, utilizando a linguagem de programação Python, com o objetivo de identificar e analisar os sentimentos que cada aluno expressa ao interagir com os colegas nesses ambientes. Os resultados obtidos, em experimentos iniciais, indicam que o minerador mostrou-se eficiente na extração das emoções dos alunos, alcançando uma acurácia de 40%. Além disso, mostraram que sentimentos positivos, como alegria e surpresa, refletem na conclusão dos MOOCs, enquanto sentimentos negativos, como tristeza e raiva, são indicativos de abandono do curso.

Palavras-chave: *MOOCs. Fóruns de discussão. Minerador de emoções. Algoritmo Naive Bayes.*

1. Introduction

The Massive Open Online Courses (MOOCs) have drawn the attention of education scholars as a new possibility of access to learning. The proposal, in general terms, is that they act as knowledge platforms for anyone, anytime, anywhere, making them an emerging and powerful learning strategy with repercussions in the technological and educational areas (Zheng et al. 2016). Zheng et al. (2016) point out that MOOCs emerged with the purpose of providing educational innovation, through pedagogical approaches that expand learning possibilities, in order to reach a large number of students.

The variety of learning resources offered by MOOCs, along with the emergence of social media, has contributed to the creation of spaces for student/teacher and student/student interaction. This interaction generates a large amount of data that evidence the learning behavior and leave traces of the educational process, which are useful for the assessment of learning (Paltoglou and Thelwall, 2012). The discussion forum, in this context, is a resource that allows participating subjects to discuss a particular topic and exchange ideas.

According to Sánchez (2005), the discussion forum for educational purposes in an online environment is defined as a communication space composed of dialogue boards, in which the written messages can be classified thematically. In these spaces, students can make contributions, refute others, clarify doubts, among others. Communication is carried out asynchronously and the messages written remain available to the participants. According to Palloff and Pratt (2004), students' interactions in discussions provide a moment of reflection on the educational content covered. The involvement in discussion forums is an important part of the activities of students who are part of the distance education modality, as the forums allow the teacher to diagnose information about the students. However, if the professor has a large number of students, the time required for him to be able to analyze the discussions will be large, possibly unfeasible for MOOCs. Thus, in order for the teacher to analyze all student responses in forums and other interaction environments, the use of computational methods can be of great value (Souza and Perry, 2019).

Given the above context, this work aimed to implement a miner for detecting emotions in texts produced by students in discussion forums. For an initial study of its applicability, a MOOC from a Brazilian platform was used. In this way, the interactions of a sample of students in the discussion forums were used to analyze whether the emotions expressed in their posts are related to the completion or abandonment of the courses. In view of this, the research question that guided this study was: Are the emotions expressed by students in MOOC discussion forums indicative of completion or abandonment? To answer this question, an emotion miner was first implemented, then it was applied to real student posts.

It is intended, with this research, in addition to providing an emotion mining tool (free of charge) for MOOC teachers/tutors, to develop an initial post analysis experiment to identify the emotions expressed by students. It should also be noted that the study presented in this article uses an emotion mining technique based on the basic emotions enunciated by Paul Ekman (1992) and that the Pycharm Integrated Development Environment and the Python programming language were used to implement the miner.

2. Related papers

This section presents some research carried out with the application of text mining techniques, to analyze discussion forums and other textual interaction environments in MOOCs, in order to recognize the emotions that are expressed by students. In the work carried out by Liu et al (2016a), for example, the authors implemented a model to perform the recognition of emotions/feelings in topics posted by students, through the mining of course comments. To do so, they automatically collected the texts of the posted comments, then used emotional topic mining, in order to identify the popularity of each course through the result of emotional recognition. The Latent Dirichlet Allocation (LDA) algorithm was used to explore possible recurring words in the texts. The authors used the model developed in the research to extract information about each student's emotions, including learning habits, preferences, styles, among others.

The research carried out by Liu et al. (2016b) also focused on sentiment analysis, whose purpose was to identify emotional and affective characteristics based on the proposition of a new model: The Multi-Swarm Particle Swarm Optimization (MSPSO). In addition to the proposition, the model was applied in an emotion recognition experiment elaborated from the selection of resources, training and, finally, the application of the test. The experimental results achieved indicated that MSPSO effectively reduced the redundancy of text features and captured discriminative features. Compared to conventional methods of selecting these features, the MSPSO performed better when selecting the same dimensions. In addition, the result of a user survey carried out by Liu et al. (2016b), indicates that 72.19% of subjects approve of the usability of recognition results and the effectiveness of resource selection.

Finally, Xing, Tang and Pei (2019) investigated the role of emotions linked to achievement in the learning experiences of students in a MOOC. The work sought to explore these emotions and their complex impact on student dropout. To do so, based on a real MOOC dataset, a miner was first implemented that automatically extracts the different emotions of achievement from student posts on forums; next, the constructed miner was applied to all posts to identify the four main realization emotions; finally, a survival modeling technique was used to quantify the effect of different student achievement emotions over the course of the course.

The results showed a different influence mechanism for emotions expressed and exposed on student survival in the MOOC, which led the authors to discuss the implications of the results in terms of projects for intervention and improvement in student retention. It should be noted that Xing, Tang and Pei (2019) based the development of their miner on classic machine learning algorithms such as Naive Bayes, Logistic Regression, Support Vector Machines and Decision Tree. Likewise, the miner developed for the present work was based on the classification technique and the Naive Bayes machine learning algorithm.

3. Emotion Mining

According to the Online Dictionary of Portuguese (2020), emotion is a moral, psychic or physical reaction, usually caused by a confusion of feelings that, in the face of some fact, situation or news, makes the body behave in accordance with this reaction, expressing respiratory, circulatory changes or commotion. According to Ekman (1992) there are six basic emotions: happiness, sadness, anger, fear, disgust and surprise. Such diversity of emotions makes, among other factors, their identification in texts a complex task. In most of the approaches developed, the aim is not to categorize emotions into specific situations and categories, but to identify them on two scales: the valence of the emotion, indicating whether the feeling is positive or negative, and the level of arousal, indicating the level of energy. associated with emotion (Thelwall, Wilkinson and Uppal, 2010).

According to Thelwall, Wilkinson, and Uppal (2010), studying emotions based on a two-dimensional scale (i.e., valence and arousal) is more reliable and provides more accurate results than studying emotions more specifically. However, with the evolution of artificial intelligence technologies, more accurate classifications can be achieved, although it is necessary to take into account that the smaller the number of classes, the greater the probability of success. As for emotion mining, several techniques have been used to automate this process. With few exceptions, these techniques are generally classified into four categories:

1. The first category, which employs keyword spotting, is based on a lexical dictionary that groups words with emotional connotations. This technique extracts the emotions of the writers by identifying these affective words from the text. For example, “happy” reflects happiness and “scared” reflects fear. These techniques are popular because of their simplicity and economic advantage (Strapparava and Valitutti, 2004);

2. The second category, which employs lexical affinity measures, is a bit more refined than keyword detection. In this technique, each word is assigned a probabilistic affinity for a certain emotion. For example, the word “success” has an 80% probability of reflecting a positive event. An example of a measure of lexical affinity is the emotional weight, used in Ma, Prendinger and Ishizuka (2005), which is calculated for each word as the proportion of emotional meanings over the total that the word can have.

3. The third category uses Natural Language Processing (NLP), a technique that employs machine learning algorithms to learn lexical affinities of words and frequencies of their occurrence, as discussed in Wilson, Wiebe and Hwa (2004).

4. The last category consists of handmade models, which use a deep understanding of the particular text to categorize emotions. Because they are complex systems, it is difficult to generalize the results to other texts. An example of such models is presented in Dyer (1987), while an improvement on artisanal models is provided by Liu, Lieberman and Selker (2003), the technique developed by which classify texts in the six basic emotions proposed by Ekman (1992).

4. Methodological procedures

The present research began with the implementation of the algorithm that constitutes the emotion mining tool for postings in discussion forums. Once ready, the miner was applied to a sample of posts from a MOOC discussion forum on a Brazilian platform. The purpose of mining emotions in courses of this type is to identify how students are feeling before they give up, so that teachers/tutors can outline intervention actions. Given this information, the research question that guides this study was formulated: Research Question – Are the emotions expressed by students in MOOC discussion forums indicative of completion or abandonment? To answer this question, some procedures were used that configure the methodology adopted in this research, which, in general terms, is summarized in: 1) Process of elaboration of the miner and 2) Application in posts of students of a MOOC. The description of these items is evidenced in sections 4.1 and 4.2.

4.1. Tool Implementation

The developed miner considers the six emotions enunciated by Ekman (1992), making just one change: the feeling of disgust was replaced by disgust, more easily identified in educational environments. The tool was built based on the union of Spotting and NLP techniques, since a base of phrases already categorized with emotions (Ekman, 1992) was used, which works as a lexical dictionary. This sentence base was used to train a machine learning algorithm, central to the NLP technique.

The technologies used to implement the tool were: 1) Python Programming Language, version 3.7; 2) PyCharm Integrated Development Environment (PyCharm Professional Edition with Anaconda plugin 2019.3.3 x64); 3) Natural Language Toolkit (NLTK) Library for Natural Language Processing in Python; and 4) Naive Bayes machine learning algorithm. For more detail on the development of this tool, the procedures performed were divided into steps that can be seen in Chart 1.

Chart 01: Stages of Emotion Miner development

ID	STAGE	DESCRIPTION
1	Generation of Training and Test Bases1	Two databases were generated in order to enable the construction of the miner, one for training and the other for testing. These bases were built with several sentences, which were classified by an expert - a psychologist -, who proposed to carry out this procedure. The Training base contains a total of 538 phrases – 112 of joy, 90 of disgust, 84 of fear, 84 of anger, 84 of surprise and 84 of sadness. The Test base has a total of 228 phrases – 45 of joy, 36 of disgust, 36 of fear, 36 of anger, 36 of suppression and 36 of sadness.
2	Removing Stop words from Bases	In a document, there are many tokens that do not have any semantic value, being useful only for the understanding and general comprehension of the text. These tokens are words classified as stop words and correspond to what is called the stop list of a Text Mining system. A list of stop words is constituted by the words that appear most in a textual mass and, normally, correspond to the articles, prepositions, punctuation, conjunctions and pronouns of a language. The identification and removal of this class of words considerably reduces the final size of the lexicon, with the beneficial consequence of increasing the performance of the system as a whole.

1 Link to access – Bases de Treinamento - Base de Teste

3	Application of the Stemming method in the Bases	**The stemming process focuses on the reduction of each word in the lexicon, until its respective root is obtained. In this way, the main benefit is the elimination of suffixes that indicate variation in the form of the word, such as plural and verb tenses. Algorithms generally don't care about using the context the word is in, and this approach doesn't seem to help much. Cases in which context helps in the stemming process are infrequent, and most words can be considered to have a single meaning.**
5	Naive Bayes Classifier Training 2	**The Naive Bayes algorithm is a supervised machine learning algorithm, so the need for the sentences belonging to the Training base to be classified. In this way, when applied to the training database, the algorithm analyzes many sentences already classified, learns the pattern embedded in these classifications and is able to generalize to new texts. It is noteworthy that the implementation of the Naive Bayes Algorithm is already encapsulated in the Python NLTK library.**
6	Naive Bayes Classifier Test	**After the algorithm is trained, it is applied to the test database. Subsequently, it performs the classifications, which are compared with those of the specialist, thus making it possible to measure their accuracy.**

Source: The author

After the implementation of the tool and its validation with the test database, an application was made to real student posts.

4.2. Application Description

The MOOC analyzed in this study is in the area of computer science, in which 894 enrollments were registered between January and May 2021. Of these enrolled students, 215 completed the course, while 609 dropped out, a rate of approximately 76% of dropout, high rates dropout rates in MOOCs are still a challenge for course managers of this type. The description of the MOOC is presented in Chart 2.

Chart 2: MOOC Description

ANALYZED MOOC	
Hours: 30 hours	Hours: 30 hours
Minimum time to obtain the certificate 3600 minutes	Minimum time to obtain the certificate 3600 minutes
Target Audience High school, college and/or graduate students	Target Audience High school, college and/or graduate students
Target Audience High school, college and/or graduate students	Target Audience High school, college and/or graduate students
modules 6	modules 6
Methodology Without tutoring	Methodology Without tutoring
Computer Science Area	Computer Science Area
Intermediate level	Intermediate level
Portuguese language	Portuguese language

Source: The author

2 The "Naive Bayes" algorithm is a probabilistic classifier based on the "Bayes Theorem", which was created by Thomas Bayes (1701 - 1761). Because it is very simple and fast, it has a relatively higher performance than other classifiers. Widely used in Machine Learning to categorize texts based on the frequency of words used.

The course has a total of four discussion forums, which constitute assessment activities that must be completed by students to obtain their certificate. In this research, posts from forum number 1 and number 3 of the MOOC were analyzed, which had the following themes: forum 1 – What is your knowledge about Web development? Have you ever produced a website or webpage? Tell us about your experience and interact with colleagues; forum 3 – Tell us about the experience of developing your first web page, share with your colleagues the positive points and your learnings, also tell us about the difficulties you encountered.

Forum 1 was chosen because it was the student's first contact with MOOC colleagues, which made it possible to measure whether, at the beginning of the course, the student would be motivated; Forum 3, in turn, was chosen because it discussed the most complex task of the course, which would be the delivery of a functional Web page, an activity that could generate many difficulties.

Five students were selected, who in total made 16 posts, which were copied to a text document and submitted to the miner's processing, which, in turn, performed the extraction of emotions from the texts. After this process, it was verified which students, among those selected, had completed the MOOC, so that, in this way, it was possible to analyze whether the feelings related to the posts were an indication of completion or abandonment of the course. The efficiencies achieved by the developed emotion miner, as well as the results of the application in the real posts, are presented in section 5.

5. Results

To simplify the understanding of the results achieved, this section is divided into two subsections: in 5.1, the effectiveness of the implemented emotion miner will be presented, while, in 5.2, the results of its application in the students' posts are exposed.

5.1. Emotion Miner Performance

The accuracy[3] of the algorithm can be seen in Figure 1, in which 93% is equivalent to the accuracy in the training base and 40% to the accuracy in the test base. It is important to emphasize that the relevant basis for the present analysis is that of tests, since, in the training basis, the algorithm knows the data, which makes the accuracy higher. Still in Figure 1, the confusion matrix generated from the test database can be visualized. In the matrix, first all the categories are printed, then the hits identified for each class of the test base.

To interpret the generated confusion matrix, it should be noted that the hits made by the algorithm are on the main diagonal. Taking joy as an example, out of a total of 48 sentences, the algorithm got 24 correct, that is, 24 sentences of joy were categorized as joy by the miner, while, on the other hand, another 24 were wrongly categorized. Among these errors, three were classified as disgust, four as fear and twelve as anger. The dot, in turn, means that no sentence was classified as sadness.

3 Proximity between the value obtained experimentally and the true value in the measurement of a physical quantity. Precision of a table or an operation.

Figure 1: Emotions Miner performance evaluation

```
Training base accuracy = 0.93122
Test base accuracy = 0.40789

CONFUSION MATRIX

          |    d          s  t |
          |  a e          u  r |
          |  l s          r  i |
          |  e g       r  p  s |
          |  g o   m   a  r  t |
          |  r s   e   i  e  e |
          |  i t   d   v  s  z |
          |  a o   o   a  a  a |
----------+--------------------+
happiness |<24> 3   4 12  5  . |
  sadness |  8<20>  1  2  3  2 |
     fear | 10  5<13>  3  2  3 |
     rage | 11  3   4<10> 3  5 |
 surprise | 18  1   .  2<12> 3 |
  sadness | 12  2   4  3  1<14>|
----------+--------------------+
(row = reference; col = test)
```

Source: The author

As mentioned, as the developed emotion miner sought to recognize six types of emotions in processed texts, there were too many labels for categorization, given that machine learning algorithms usually classify data into two or three categories. In this sense, although the accuracy of the algorithm tends to appear low, it is considerable due to the number of categories that must be recognized.

To validate this statement, two situations must be taken into account: 1) random classification and 2) classification according to the class that has the highest number of elements. In the first one, the percentage of correct categorization of a new sentence by drawing must be verified, which corresponds to 16% (100/6); in the second, the percentage of correct answers is analyzed if all new sentences were labeled as the category that has more elements, in the case of joy (48 sentences in the test base out of a total of 228), which is equivalent to 21% (48/228). Thus, when analyzing these parameters, it can be considered that the developed miner has satisfactory results regarding the process of identifying emotions.

5.2. Application Results

As explained in subsection 4.2 of the application description, the posts of five students in forums 1 and 3 were analyzed, who wrote 16 posts, which were submitted to the miner's processing and resulted in the classification shown in Chart 3.

Chart 3: Result of the Application of the Emotions Miner

Student	Forum	Posts*	Miner's Result
1	1	Hello everyone, everything good? I've never produced any website so I hope you learn everything you need to in this course.	**happiness** happiness: 0.625408 / heartbreak: 0.076334 fear: 0.012036 / rage: 0.037552 surprise: 0.197367 / sadness: 0.051302

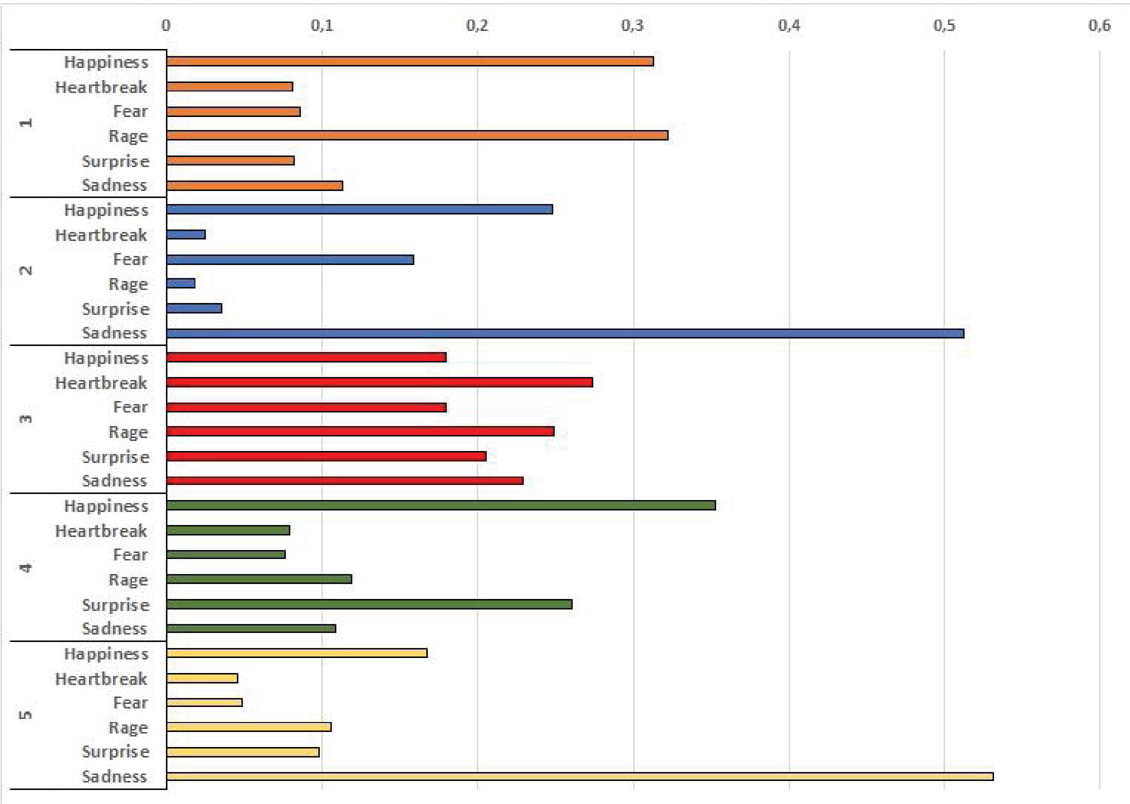
2	1	Good afternoon, I want to learn how to make websites in this course.	**sadness** happiness: 0.308301 / heartbreak: 0.020337 fear: 0.103571 / rage: 0.020277 surprise: 0.064755 / sadness: 0.482759
3	1	I don't know how to make websites or web pages, but I want to learn in this course!	**sadness** happiness: 0.308301 / heartbreak: 0.020337 fear: 0.103571 / rage: 0.020277 surprise: 0.064755 / sadness: 0.482759
4	1	Hello, I have some experience in website development, because I took the technical course in computer science, so I would like to improve and learn new things.	**happiness** happiness: 0.431392 / heartbreak: 0.104940 fear: 0.023412 / rage: 0.071261 surprise: 0.227579 / sadness: 0.141416
5	1	Good morning colleagues, how are you? I've made websites, but only with tools like WIX, so I want to learn how creating web pages really works.	**sadness** happiness: 0.060954 / heartbreak: 0.004633 fear: 0.004007 / rage: 0.022501 surprise: 0.004226 / sadness: 0.903679
1	3	I found it very difficult to make the site, I didn't get it on the first attempt, but little by little I was doing it.	**happiness** happiness: 0.276075 / heartbreak: 0.115748 fear: 0.174988 / rage: 0.222804 surprise: 0.025429 / sadness: 0.184955
1	3	Yes me too, I didn't get everything I needed with this course, to finish the task I had to watch classes on YouTube.	**rage** happiness: 0.038674 / heartbreak: 0.051483 fear: 0.070128 / rage: 0.708679 surprise: 0.025992 / sadness: 0.105044
2	3	I found it easy to do, but I didn't really like the result.	**sadness** happiness: 0.188811 / heartbreak: 0.029723 fear: 0.214864 / rage: 0.017596 surprise: 0.006097 / sadness: 0.542909
3	3	I found the task very difficult, it took me a while to finish it. My main difficulty was inserting figures and leaving them where I wanted them.	**rage** happiness: 0.089446 / heartbreak: 0.037502 fear: 0.172122 / rage: 0.523910 surprise: 0.075936 / sadness: 0.101084
3	3	Also, I found the course did not explain everything that was needed.	**rage** happiness: 0.126045 / heartbreak: 0.178579 fear: 0.190967 / rage: 0.448544 surprise: 0.016451 / sadness: 0.039413
3	3	I would need more classes, to improve.	**fear** happiness: 0.298963 / heartbreak: 0.067842 fear: 0.322595 / rage: 0.191737 surprise: 0.066436 / sadness: 0.052427
3	3	I had a lot of doubts about the tables and figures.	**sadness** happiness: 0.076945 / heartbreak: 0.065843 fear: 0.110539 / rage: 0.065700 surprise: 0.209820 / sadness: 0.471152
4	3	The task was simple, I had no difficulties in doing it, and the page was pretty cool.	**surprise** happiness: 0.274905 / heartbreak: 0.055009 fear: 0.130879 / rage: 0.166642 surprise: 0.295703 / sadness: 0.076863

5	3	Good morning, I found the task easy and I learned how to make the page from the beginning, so I liked the activity.	**sadness** happiness: 0.010421 / heartbreak: 0.039144 fear: 0.020893 / rage: 0.080763 surprise: 0.027984 / sadness: 0.820795
5	3	I didn't have the same difficulties, but I also searched the internet for some doubts.	**surprise** happiness: 0.118524 / heartbreak: 0.110139 fear: 0.095486 / rage: 0.290642 surprise: 0.305738 / sadness: 0.079471
5	3	I liked the result it was very beautiful!	**happiness** happiness: 0.481508 / heartbreak: 0.031832 fear: 0.075734 / rage: 0.031762 surprise: 0.056362 / sadness: 0.322801

* Portuguese errors and abbreviated words have been fixed so that the posts are more readable
Source: The author

For a better understanding of the information contained in Chart 3, a graph was prepared (Figure 2) that presents the predominant emotion in the posts of each of the students, for this a simple average was applied to the results presented by the miner for the publications of the students, thus way it is possible to visualize in general how they were feeling. For example, student 4 made two posts, so the values obtained by the miner for each emotion were added and then divided by two.

Figure 2: Average Miner's Result per Student



Source: The author

As can be seen in Chart 3, the algorithm, in addition to indicating the predominant emotion, also indicates the probability that the text belongs to other categories. For a better understanding, consider the first post of student 1 and its classification: despite surprise having a high probability, around 20%, joy had a percentage above 50%, more precisely 62%, which brings some security in the classification. However, in the third post of student 5 (penultimate line of the chart), which was classified as surprise, it is difficult to say that this is really the emotion expressed, as surprise has a 30% probability, while anger appears with 29%. In this case, the identification has to take into account the two categories that had the highest percentages. After processing the posts and determining the predominant emotions in each of them, an analysis of the students' completion was carried out, in which it was possible to verify that students 1, 4 and 5 completed the MOOC and students 2 and 3 did not. concluded it. The synthesized analysis can be seen in Chart 4.

Chart 4: Summary of Analysis

Student	Post 1	Post 2	Post 3	Post 4	Post 5	STATUS
1	Happiness	Happiness	Rage			concluded
2	Sadness	Sadness				abandoned
3	Sadness	Rage	Rage	Fear	Sadness	abandoned
4	Happiness	Surprise				concluded
5	Sadness	Sadness	Surprise	Happiness		concluded

Source: The author

As can be seen in Chart 4, there is an indication that the preponderance of sadness and anger cause abandonment, and that surprise and joy indicate the conclusion. However, the other emotions with high probabilities, in the classification, should also be evaluated, and more students investigated. In view of this, more studies need to be developed so that these results can be generalized, given that this is a preliminary study, which presents some evidence with the potential to be validated and analyzed in more depth, especially considering a greater number of posts, the main element to make it possible to generalize the results presented in this investigation. However, although the results are still superficial, it is fair to say that the emotion miner can be an interesting tool for teachers/tutors to be able to identify students who are unmotivated before giving up and can plan actions for them to complete the course. MOOC.

6. Conclusions

The present work had as main objective the implementation of an emotion miner for postings on MOOCs discussion forums, and, to validate its applicability, a preliminary study was carried out with the posts of five students of a MOOC course on a Brazilian platform. The study also sought to identify which positive and/or negative feelings extracted from the students' posts reflect on the completion or abandonment of the courses.

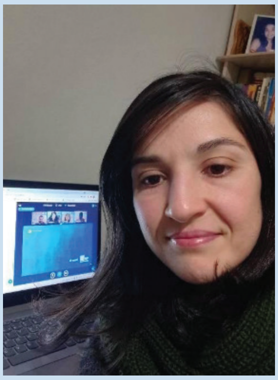
It was possible, with the conduction of this research, to identify that the miner presents an acceptable accuracy in the classification of student posts, around 40%, for the 6 specified emotions. It was also possible to identify that, if the students' posts are categorized with emotions of sadness and anger, these are indications that the student may drop out of the course, and, if they are classified with emotions of joy or surprise, these are indicative that the student may abandon the course. student will finish the course.

As future works, we intend to implement an interface so that MOOC teachers/tutors can load all posts from a forum and the miner processes the predominant emotion of each student during the forum. In addition, it is expected to carry out studies with more students, so that it is possible to validate the preliminary results identified in this research.

Thanks

The present work was carried out with the support of the Federal Institute of Education, Science and Technology of Rio Grande do Sul.

Biodata and Author Contact



SOUZA, V. F. is a doctoral student at the Graduate Program in Informatics in Education (PPGIE) at the Federal University of Rio Grande do Sul (UFRGS). Master in Informatics by the Graduate Program in Informatics (PPGI) of the Federal Technological University of Paraná (UTFPR), in the area of Applied Computing, with emphasis on Software Engineering. He has a Specialization in Inclusive Special Education, with an emphasis on Assistive Technologies from the State University of Northern Paraná (UENP). She is graduated in Information Systems also by UENP - Bachelor in Information Systems and Licentiate in Computing. He also completed a Degree in Mathematics in the Pedagogical Training Program at UTFPR. He is currently an exclusive professor at the Federal Institute of Education, Science and Technology of Rio Grande do Sul (IFRS) at Campus Ibirubá, and is on leave to pursue his Doctorate. He teaches classes in Computer Science Courses, Integrated Computer Technician in High School, Mathematics Degree and Specialization in Language Teaching and its Technologies. He is a member of the Alto Jacuí Interdisciplinary Computing Research Group. Interested in Educational Data Mining, Learning Analytics, Artificial Intelligence - Machine Learning and Deep Leaning, Digital Systems and Educational Robotics.

ORCID: https://orcid.org/0000-0002-8804-2680

E-MAIL: *vanessa.souza@ibiruba.ifrs.edu.br

References

DICIONÁRIO ONLINE. 2020. Available at: https://www.dicio.com.br/ . Access in: april , 2020.

EKMAN, P. An argument for basic emotions. **Cognition & Emotion,** v. 6, n. 3, p. 169-200, 1992.

LIU, H.; LIEBERMAN, H.; SELKER, T. A Model of Textual Affect Sensing Using Real-World Knowledge. In: **Proc. Eighth Int'l Conf. Intelligent User Interfaces,** 2003, p.369-374.

LIU, Z. *et al*. Sentiment recognition of online course reviews using multi-swarm optimization-based selected feature. **Neurocomputing,** v. 17, n. 12, p. 11-20, 2016b.

LIU, Z. *et al.* Emotion and Associated Topic Detection for Course Comments in a MOOC Platform. In: **International Conference on Educational Innovation through Technology (EITT),** Taiwan, China, 2016a, p.15-19.

MA, C.; PRENDINGER, H.; ISHIZUKA, M. Emotion Estimation and Reasoning Based on Affective Textual Interaction. In: **First Int'l Conf. Affective Computing and Intelligent Interaction (ACII '05),** 2005, p. 622-628.

PALLOFF, R. M.; PRATT, K. O aluno virtual: um guia para trabalhar com estudantes on-line. Porto Alegre: **Artmed,** 2004.

PALTOGLOU G.; THELWALL, M. Twitter, MySpace, Digg: Unsupervised Sentiment Analysis. **Social Media, in ACM Transactions on Intelligent Systems and Technology,** vol. 3, no. 4, 2012.

SÁNCHEZ, L. P. El foro virtual como espacio educativo: propuestas didácticas para su uso. **Verista Quaderns Digitals Net**, n.40, p.1-18, 2005. Disponível em: http://www.quadernsdigitals.net/datos_web/hemeroteca/r_1/nr_662/a_8878/8878.htm Acesso em: 27 mar. 2019.

SOUZA, V. F.; PERRY, G. T. Mineração de Texto em MOOCs: Análise da Relevância Temática de Postagens em Fóruns de Discussão. **RENOTE - Revista Novas Tecnologias na Educação**. Porto Alegre, v. 17, n. 3, p. 204-2013, 2019.

STRAPPARAVA, C.; VALITUTTI, A. WordNet-Affect: an affective extension of WordNet. In: **4th International Conference on Language Resources and Evaluation (LREC '04)**, Lisbon, Portugal, 2004, p. 1083–1086.

THELWALL, M.; WILKINSON, D.; UPPAL, S. Data mining emotion in social network communication: Gender differences in MySpace. **Journal of the American Society for Information Science and Technology**, v. 61, n. 1, p. 190-199, 2010.

XING, W.; TANG, H.; PEI, B. Beyond positive and negative emotions: Looking into the role of achievement emotions in discussion forums of MOOCs. **The Internet and Higher Education,** v. 43, p. 1-9, 2019.

WILSON, T.; WIEBE, J.; HWA, R. Just how mad are you? Finding strong and weak opinion clauses. In: **21st Conference of the American Association for Artificial Intelligence,** San Jose, US, 2004, p. 761–769.

ZHENG, S. *et al*. Ask the Instructors: Motivations and Challenges of Teaching Massive Open Online Courses. In: **19th Conference on Computer Supported Cooperative Work & Social Computing**, 2016, p. 206-221.