

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

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

ISSN 2177-8310
DOI: 10.18264/eadf.v12i2.1732

Vanessa Faria de Souza^{1*}

¹ Instituto Federal de Educação, Ciência e Tecnologia do Rio Grande do Sul – Campus Ibirubá – Ibirubá – RS – Brasil.

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

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.



Recebido 25/01/2022
Aceito 06/06/2022
Publicado 07/06/2022

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>

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

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.

1. Introdução

Os *Massive Open Online Courses* (MOOCs) têm chamado a atenção dos estudiosos da educação como uma nova possibilidade de acesso à aprendizagem. A proposta, em termos gerais, é que atuem como plataformas de conhecimento para qualquer um, a qualquer hora, em qualquer lugar, tornando-as uma emergente e poderosa estratégia de aprendizagem com repercussão nas áreas tecnológica e educacional (Zheng *et al.* 2016). Zheng *et al.* (2016) salientam que os MOOCs despontaram com o propósito de proporcionar inovação educacional, por meio de abordagens pedagógicas que expandem as possibilidades de aprendizagem, de forma a atingir uma grande quantidade de alunos.

A variedade de recursos de aprendizagem oferecidos pelos MOOCs, juntamente com o surgimento das mídias sociais, contribuiu com a criação de espaços para a interação aluno/professor e aluno/aluno. Essa interação gera uma grande quantidade de dados que evidenciam o comportamento de aprendizagem e deixam rastros do processo educacional, os quais são úteis para a avaliação da aprendizagem (Paltoglou e Thelwall, 2012). O fórum de discussão, nesse contexto, é um recurso que permite aos sujeitos participantes discutir um determinado assunto e trocar ideias.

Segundo Sánchez (2005), o fórum de discussão para fins educacionais em um ambiente online é definido como um espaço de comunicação composto por quadros de diálogo, nos quais as mensagens redigidas podem ser classificadas tematicamente. Nesses espaços, os alunos podem realizar contribuições, refutar outras, esclarecer dúvidas, entre outros. A comunicação é efetuada de forma assíncrona e as mensagens redigidas permanecem à disposição dos participantes. De acordo com Palloff e Pratt (2004), as interações dos alunos nas discussões proporcionam um momento de reflexão sobre os conteúdos educacionais abordados. O envolvimento em fóruns de discussão é uma parte importante das atividades dos alunos que integram a modalidade de educação a distância, pois os fóruns permitem que o professor possa diagnosticar informações sobre os discentes. No entanto, se o docente tiver grande quantidade

de alunos, o tempo necessário para que ele consiga analisar as discussões será grande, possivelmente inviável para MOOCs. Dessa forma, a fim de que o professor possa analisar todas as respostas dos alunos em fóruns e demais ambientes de interação, o emprego de métodos computacionais pode ser de grande valia (Souza e Perry, 2019).

Diante do contexto exposto, este trabalho teve como objetivo realizar a implementação de um minerador para detecção de emoções em textos produzidos pelos alunos em fóruns de discussão. Para um estudo inicial de sua aplicabilidade, foi utilizado um MOOC de uma plataforma brasileira. Dessa forma, as interações de uma amostra de alunos nos fóruns de discussão foram utilizadas para analisar se as emoções expressas em suas postagens têm relação com a conclusão ou com o abandono dos cursos. À vista disso, a questão de pesquisa que norteou este estudo foi: As emoções expressas por alunos em fóruns de discussão de MOOCs são indicativas de conclusão ou abandono? Para responder tal questionamento primeiramente foi implementado um minerador de emoções, em seguida o mesmo foi aplicado a postagens reais de alunos.

Pretende-se, com esta pesquisa, além de disponibilizar uma ferramenta de mineração de emoções (de forma gratuita e livre) para professores/tutores de MOOCs, desenvolver um experimento inicial de análise de postagens para identificar as emoções expressas pelos alunos. Cabe ainda salientar que o estudo apresentado neste artigo utiliza técnica de mineração de emoções baseada nas emoções básicas enunciadas por Paul Ekman (1992) e que foram utilizados o Ambiente de Desenvolvimento Integrado Pycharm e a linguagem de programação Python, para a implementação do minerador.

2. Trabalhos Relacionados

Esta seção apresenta algumas pesquisas realizadas com a aplicação de técnicas de mineração de textos, para analisar fóruns de discussão e demais ambientes de interação textual em MOOCs, com o intuito de reconhecer as emoções que são expressas pelos alunos. No trabalho realizado por Liu *et al.* (2016a), por exemplo, os autores implementaram um modelo para realizar o reconhecimento de emoções/sentimentos em tópicos postados pelos alunos, por meio da mineração de comentários do curso. Para tanto, coletaram automaticamente os textos dos comentários postados, usando, em seguida, a mineração de tópicos emocionais, com o objetivo de identificar a popularidade de cada curso por meio do resultado do reconhecimento emocional. O algoritmo de Latent Dirichlet Allocation (LDA) foi usado para explorar as possíveis palavras recorrentes nos textos. Os autores empregaram o modelo desenvolvido na pesquisa para extrair informações sobre as emoções de cada aluno, incluindo hábitos de aprendizagem, preferências, estilos, dentre outros.

A pesquisa realizada por Liu *et al.* (2016b) também teve como cerne a análise de sentimentos, cujo propósito foi identificar características emocionais e afetivas a partir da proposição de um novo modelo: o Multi-Swarm Particle Swarm Optimization (MSPSO). Além da proposição, o modelo foi aplicado em um experimento de reconhecimento de emoções elaborado a partir da seleção de recursos, do treinamento e, por fim, da aplicação do teste. Os resultados experimentais atingidos indicaram que MSPSO efetivamente reduziu a redundância de recursos de texto e capturou recursos discriminativos. Comparado com os métodos convencionais de seleção desses recursos, o MSPSO obteve o melhor desempenho ao selecionar as mesmas dimensões. Além disso, o resultado de uma pesquisa com usuários, realizada por Liu *et al.* (2016b), indica que 72,19% dos sujeitos aprovam a usabilidade dos resultados de reconhecimento e a eficácia da seleção dos recursos.

Por fim, Xing, Tang e Pei (2019) investigaram o papel das emoções vinculadas à conquista nas experiências de aprendizagem dos alunos de um MOOC. O trabalho procurou explorar essas emoções e seu impacto complexo na evasão de estudantes. Para tanto, baseando-se em um conjunto de dados reais de um MOOC, primeiramente foi implementado um minerador que extraia automaticamente as diferentes emoções de conquista das postagens dos alunos em fóruns; na sequência, o minerador construído foi aplicado a todas as postagens para identificar as quatro emoções de realização principais; por fim, uma

técnica de modelagem de sobrevivência foi usada para quantificar o efeito de diferentes emoções de conquista do aluno no decorrer do curso.

Os resultados mostraram um mecanismo de influência diferente para as emoções expressas e expostas na sobrevivência do aluno no MOOC, o que levou os autores a discutir as implicações dos resultados em termos de projetos para intervenção e melhoramento na retenção de estudantes. Cabe destacar que Xing, Tang e Pei (2019) basearam o desenvolvimento de seu minerador em algoritmos de aprendizagem de máquina clássicos como Naive Bayes, Logistic Regression, Suport Vector Machines e Decision Tree. Da mesma forma, o minerador desenvolvido para o presente trabalho foi baseado na técnica de classificação e no algoritmo de aprendizagem de máquina Naive Bayes.

3. Mineração de Emoções

De acordo com o Dicionário Online de Português (2020), emoção é uma reação moral, psíquica ou física, geralmente causada por uma confusão de sentimentos que, diante de algum fato, situação ou notícia, faz com que o corpo se comporte de acordo com essa reação, expressando alterações respiratórias, circulatórias ou comoção. Segundo Ekman (1992) existem seis emoções básicas: felicidade, tristeza, raiva, medo, nojo e surpresa. Tal diversidade de emoções torna, entre outros fatores, sua identificação em textos uma tarefa complexa. Na maioria das abordagens desenvolvidas, não se busca categorizar emoções em situações e categorias específicas, mas identificá-las em duas escalas: a valência da emoção, indicando se o sentimento é positivo ou negativo, e o nível de excitação, indicando o nível de energia associado com a emoção (Thelwall, Wilkinson e Uppal, 2010).

De acordo com Thelwall, Wilkinson e Uppal (2010), estudar emoções com base em uma escala bidimensional (ou seja, valência e excitação) é mais confiável e fornece mais resultados precisos do que estudar emoções de forma mais especificada. Todavia, com a evolução das tecnologias de inteligência artificial, classificações mais precisas podem ser alcançadas, embora seja necessário levar em consideração que, quanto menor o número de classes, maior a probabilidade de acerto. Quanto à mineração de emoções, várias técnicas têm sido usadas para automatizar esse processo. Com poucas exceções, essas técnicas são geralmente classificadas em quatro categorias:

1. A primeira categoria, que emprega Spotting de palavras-chave, é baseada em um dicionário léxico que agrupa palavras dotadas de conotações de emoções. Essa técnica extrai as emoções dos escritores, identificando essas palavras afetivas do texto. Por exemplo, “feliz” reflete felicidade e “assustado” reflete medo. Essas técnicas são populares devido à sua simplicidade e vantagem econômica (Strapparava e Valitutti, 2004);
2. A segunda categoria, que emprega medidas de afinidade lexical, é um pouco mais refinada do que a detecção de palavras-chave. Nessa técnica, a cada palavra é atribuída uma afinidade probabilística para uma certa emoção. Por exemplo, a palavra “sucesso” tem 80% de probabilidade de refletir um evento positivo. Um exemplo de medida de afinidade lexical é o peso emocional, usada em Ma, Prendinger e Ishizuka (2005), o qual é calculado para cada palavra como a proporção de sentidos emocionais sobre o total que a palavra pode ter.
3. A terceira categoria usa o Processamento de Linguagem Natural (PLN), técnica que emprega algoritmos de aprendizagem de máquina para aprender afinidades lexicais das palavras e frequências de suas ocorrências, como discutido em Wilson, Wiebe e Hwa (2004).
4. A última categoria consiste em modelos artesanais, os quais usam uma compreensão profunda do texto em particular para categorizar as emoções. Por serem sistemas complexos, é difícil generalizar os resultados para outros textos. Um exemplo de tais modelos é apresentado em Dyer (1987), enquanto uma melhoria nos modelos artesanais é fornecida por Liu, Lieberman e Selker (2003), a técnica desenvolvida pelos quais classifica os textos nas seis emoções básicas propostas por Ekman (1992).

4. Procedimentos Metodológicos

A presente pesquisa teve início com a implementação do algoritmo que constitui a ferramenta de mineração de emoções para postagens em fóruns de discussão. Depois de pronto, o minerador foi aplicado a uma amostra de postagens de um fórum de discussão de um MOOC de uma plataforma brasileira. O intuito de minerar as emoções em cursos desse tipo é identificar como os alunos estão sentindo-se antes de desistirem, para que professores/tutores possam traçar ações de intervenção. Diante dessa informação, formulou-se a questão de pesquisa que norteia este estudo: Questão de Pesquisa – As emoções expressas por alunos em fóruns de discussão de MOOCs são indicativas de conclusão ou abandono? Para responder essa questão, foram empregados alguns procedimentos que configuram a metodologia adotada nesta pesquisa, que, em termos gerais, resume-se em: 1) Processo de elaboração do minerador e 2) Aplicação em postagens de alunos de um MOOC. A descrição desses itens é evidenciada nas seções 4.1 e 4.2.

4.1. Implementação da Ferramenta

O minerador desenvolvido considera as seis emoções enunciadas por Ekman (1992), fazendo apenas uma alteração: o sentimento de nojo foi substituído por desgosto, mais facilmente identificado em ambientes educacionais. A ferramenta foi construída baseando-se na união das técnicas de Spotting e PLN, dado que foi utilizada uma base de frases já categorizadas com as emoções (Ekman, 1992) que funciona como o dicionário léxico. Essa base de frases foi utilizada para o treinamento de um algoritmo de aprendizagem de máquina, central à técnica de PLN.

As tecnologias utilizadas para a implementação da ferramenta foram: 1) Linguagem de Programação Python, versão 3.7; 2) Ambiente de Desenvolvimento Integrado PyCharm (PyCharm Professional Edition with Anaconda plugin 2019.3.3 x64); 3) Biblioteca Natural Language Toolkit (NLTK¹) para Processamento de Linguagem Natural em Python; e 4) Algoritmo de aprendizagem de máquina Naive Bayes. Para maior detalhamento do desenvolvimento dessa ferramenta, os procedimentos realizados foram divididos em etapas que podem ser visualizadas no Quadro 1.

Quadro 01: Etapas do desenvolvimento do Minerador de Emoções

ID	ETAPA	DESCRIÇÃO
1	Geração das Bases de Treinamento e Teste ²	Duas bases de dados foram geradas, a fim de viabilizar a construção do minerador, uma para o treinamento e outra para o teste. Essas bases foram construídas com várias frases, as quais foram classificadas por um especialista – uma psicóloga –, que se propôs a realizar esse procedimento. A base de <i>Treinamento</i> contém um total de 538 frases – 112 de alegria, 90 de desgosto, 84 de medo, 84 de raiva, 84 de surpresa e 84 de tristeza. A base de <i>Teste</i> possui um total de 228 frases – 45 de alegria, 36 de desgosto, 36 de medo, 36 de raiva, 36 de surpresa e 36 de tristeza.
2	Remoção de <i>Stopwords</i> das Bases	Em um documento, existem muitos tokens que não apresentam nenhum valor semântico, sendo úteis apenas para o entendimento e a compreensão geral do texto. Esses tokens são palavras classificadas como <i>stopwords</i> e correspondem ao que é chamado de <i>stoplist</i> de um sistema de Mineração de Textos. Uma lista de <i>stopwords</i> é constituída pelas palavras de maior aparição em uma massa textual e, normalmente, correspondem aos artigos, preposições, pontuação, conjunções e pronomes de uma língua. A identificação e remoção dessa classe de palavras reduz de forma considerável o tamanho final do léxico, tendo como consequência benéfica o aumento de desempenho do sistema como um todo.

¹ O NLTK é uma biblioteca da linguagem Python para Processamento de Linguagem Natural e *Text Analytics*, que foi originalmente criada para o ensino de PLN, mas que vem sendo amplamente adotada no desenvolvimento de aplicações de PLN em geral.

² Link para acesso – Bases de Treinamento - Base de Teste

ID	ETAPA	DESCRIÇÃO
3	Aplicação do método de <i>Stemming</i> nas Bases	O processo de <i>stemming</i> concentra-se na redução de cada palavra do léxico, até que seja obtida sua respectiva raiz. Desta maneira, tem-se como principal benefício a eliminação de sufixos que indicam variação na forma da palavra, como plural e tempos verbais. Os algoritmos em geral não se preocupam com o uso do contexto no qual a palavra se encontra, e essa abordagem parece não ajudar muito. Casos em que o contexto ajuda no processo de <i>stemming</i> não são frequentes, e a maioria das palavras pode ser considerada como apresentando um significado único.
5	Treinamento do Classificador Naive Bayes ³	O algoritmo Naive Bayes é um algoritmo de aprendizagem de máquina supervisionado, por isso a necessidade das frases pertencentes a base de Treinamento estarem classificadas. Dessa forma, ao ser aplicado sobre a base de dados de treinamento, o algoritmo analisa muitas frases já classificadas, aprende o padrão embutido nessas classificações e consegue generalizar para novos textos. Destaca-se que a implementação do Algoritmo Naive Bayes já vem encapsulada na biblioteca NLTK do Python.
6	Teste do Classificador Naive Bayes	Após o algoritmo ser treinado, é aplicado sobre a base de dados de teste. Na sequência, realiza as classificações, que são comparadas com as do especialista, tornando-se possível, dessa maneira, medir sua acurácia.

Fonte: Autor

Após a implementação da ferramenta e sua validação com a base de dados de teste, foi realizada uma aplicação a postagens reais de alunos.

4.2. Descrição da Aplicação

O MOOC analisado neste estudo é da área da ciência da computação, no qual foram registradas 894 matrículas entre janeiro e maio de 2021. Desses alunos matriculados, 215 concluíram o curso, enquanto 609 desistiram, um índice de aproximadamente 76% de desistência, altos índices de desistência em MOOCs ainda constituem um desafio para os gestores de cursos deste tipo. A descrição do MOOC é apresentada no Quadro 2.

Quadro 2: Descrição do MOOC

MOOC ANALISADO	
Carga-horária:	30 horas
Prazo mínimo para obtenção do certificado	3600 minutos
Público-alvo	Estudantes do ensino médio, superior e/ou Pós-Graduação
Pré-Requisitos	Conhecimento básico em informática e internet
Módulos	6
Metodologia	Sem tutoria
Área	Ciência da Computação
Nível	Intermediário
Idioma	Português

Fonte: Autor

³ O algoritmo "Naive Bayes" é um classificador probabilístico baseado no "Teorema de Bayes", o qual foi criado por Thomas Bayes (1701 - 1761). Por ser muito simples e rápido, possui um desempenho relativamente maior do que outros classificadores. Muito utilizado na área de Aprendizado de Máquina para categorizar textos com base na frequência das palavras usadas.

O curso tem um total de quatro fóruns de discussão, os quais constituem atividades avaliativas que devem ser completadas pelos alunos para obter seu certificado. Nesta pesquisa, foram analisadas postagens do fórum número 1 e número 3 do MOOC, que tinham os seguintes temas: *fórum 1* – Qual o seu conhecimento sobre desenvolvimento Web? Você já produziu algum site ou página Web? Conte como foi sua experiência e interaja com os colegas; *fórum 3* – Conte como foi a experiência de desenvolvimento de sua primeira página Web, compartilhe com os colegas os pontos positivos e suas aprendizagens, também conte sobre as dificuldades que encontrou.

O fórum 1 foi escolhido por tratar-se do primeiro contato do aluno com os colegas do MOOC, o que possibilitou medir se, no início do curso, o estudante estaria motivado; o fórum 3, por sua vez, foi escolhido por tratar-se da discussão da tarefa mais complexa do curso, que seria a entrega de uma página Web funcional, atividade que poderia gerar muitas dificuldades.

Foram selecionados cinco alunos, que no total realizaram 16 postagens, as quais foram copiadas para um documento de texto e submetidas ao processamento do minerador, que, por sua vez, realizou a extração das emoções dos textos. Após esse processo, foi verificado quais alunos, dentre os selecionados, haviam concluído o MOOC, para que, dessa forma, fosse possível analisar se os sentimentos relacionados com as postagens eram um indicativo de conclusão ou abandono do curso. A eficácia alcançada pelo minerador de emoções desenvolvido, assim como os resultados da aplicação nas postagens reais, são apresentados na seção 5.

5. Resultados

Para simplificar o entendimento dos resultados alcançados, esta seção está dividida em duas subseções: em 5.1, será apresentada a eficácia do minerador de emoções implementado, enquanto, em 5.2, são expostos os resultados de sua aplicação nas postagens dos alunos.

5.1. Desempenho do Minerador de Emoções

A acurácia⁴ do algoritmo pode ser visualizada na Figura 1, na qual os 93% equivalem à acurácia na base de treinamento e os 40% à acurácia na base de teste. É importante ressaltar que a base relevante para a presente análise é a de testes, já que, na base de treinamento, o algoritmo conhece os dados, o que faz com que a acurácia seja mais alta. Ainda na Figura 1, pode ser visualizada a matriz de confusão gerada a partir da base de dados de teste. Na matriz, primeiramente são impressas todas as categorias, depois os acertos identificados para cada classe da base de teste.

Para interpretar a matriz de confusão gerada, deve-se observar que os acertos efetuados pelo algoritmo constam na diagonal principal. Tomando a alegria como exemplo, de um total de 48 frases, o algoritmo acertou 24, ou seja, 24 frases de alegria foram categorizadas como alegria pelo minerador, enquanto, por outro lado, outras 24 foram categorizadas de forma errônea. Dentre esses erros, três foram classificados como desgosto, quatro como medo e doze como raiva. O ponto, por sua vez, significa que nenhuma frase foi classificada como tristeza.

⁴ Proximidade entre o valor obtido experimentalmente e o valor verdadeiro na medição de uma grandeza física. Precisão de uma tabela ou de uma operação.

Figura 1: Avaliação do desempenho do Minerador de Emoções

```
Acurácia na base de treinamento = 0.9312267657992565
Acurácia na base de teste = 0.40789473684210525
MATRIZ DE CONFUSÃO
|   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 |
-----+-----+
alegria |<24> 3 4 12 5 . |
desgosto | 8<20> 1 2 3 2 |
medo | 10 5<13> 3 2 3 |
raiva | 11 3 4<10> 3 5 |
surpresa | 18 1 . 2<12> 3 |
tristeza | 12 2 4 3 1<14>|
-----+-----+
(row = reference; col = test)
```

Fonte: Autor

Conforme mencionado, como o minerador de emoções desenvolvido buscou reconhecer seis tipos de emoções nos textos processados, havia muitos rótulos para categorização, dado que geralmente algoritmos de aprendizagem de máquina classificam os dados em duas ou três categorias. Nesse sentido, embora a acurácia do algoritmo tenda a parecer baixa, essa é considerável devido ao número de categorias que devem ser reconhecidas.

Para validação dessa afirmação, deve-se levar em conta duas situações: 1) a classificação aleatória e 2) a classificação de acordo com a classe que apresenta maior quantidade de elementos. Na primeira, deve-se verificar a porcentagem de acerto da categorização de uma nova frase por sorteio, que corresponde a 16% (100/6); na segunda, analisa-se qual seria a porcentagem de acerto se todas as novas frases fossem rotuladas como a categoria que possui mais elementos, no caso alegria (48 frases na base de teste de um total de 228), que equivale a 21% (48/228). Dessa forma, ao analisar esses parâmetros, pode-se considerar que o minerador desenvolvido tem resultados satisfatórios quanto ao processo de identificação das emoções.

5.2. Resultados da Aplicação

Como exposto na subseção 4.2 da descrição da aplicação, foram analisadas as postagens de cinco alunos nos fóruns 1 e 3, os quais redigiram 16 postagens, que foram submetidas ao processamento do minerador e resultaram na classificação exposta no Quadro 3.

Quadro 3: Resultado da Aplicação do Minerador de Emoções

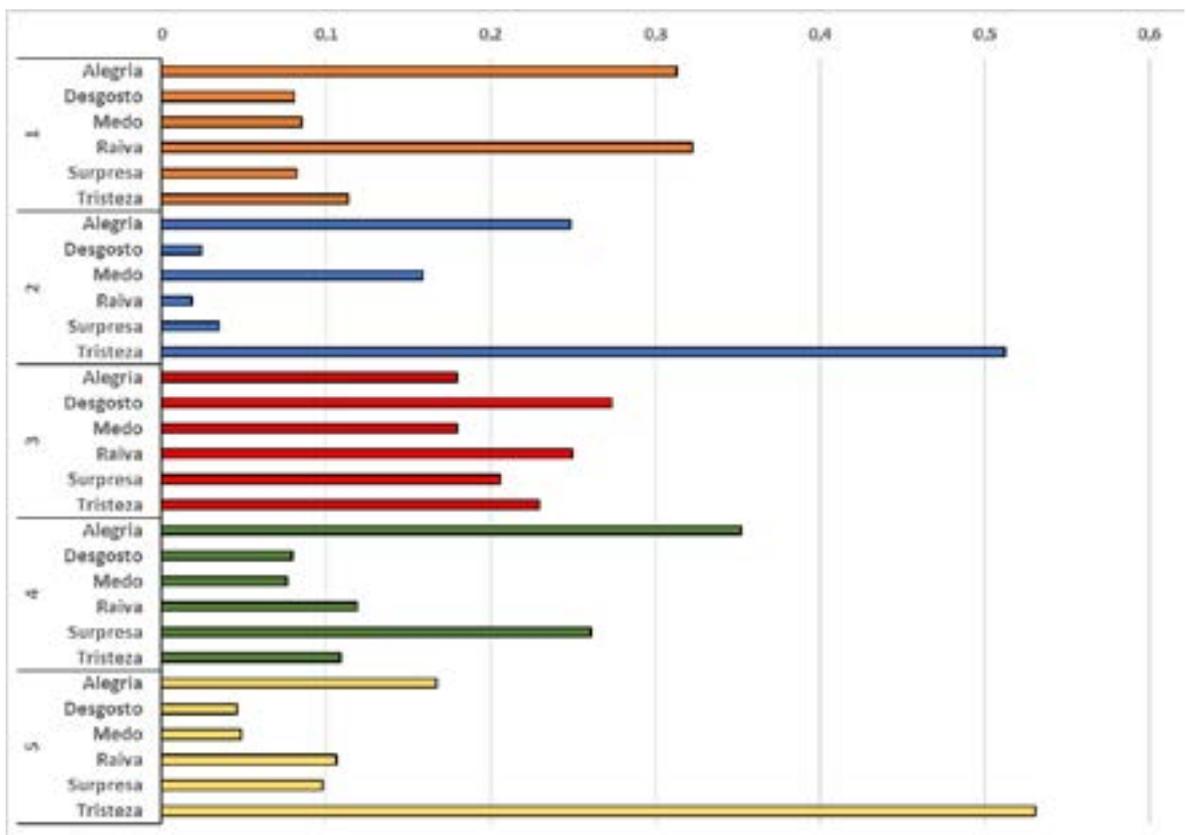
Aluno	Fórum	Postagens*	Resultado do Minerador
1	1	Olá Pessoal, tudo bem? Nunca produzi nenhum site espero que aprenda tudo que for necessário neste curso.	alegria alegria: 0.625408 / desgosto: 0.076334 medo: 0.012036 / raiva: 0.037552 surpresa: 0.197367 / tristeza: 0.051302
2	1	Boa tarde, quero aprender a fazer sites nesse curso.	tristeza alegria: 0.308301 / desgosto: 0.020337 medo: 0.103571 / raiva: 0.020277 surpresa: 0.064755 / tristeza: 0.482759
3	1	Não sei fazer sites nem páginas Web, mas quero aprender nesse curso!	tristeza alegria: 0.308301 / desgosto: 0.020337 medo: 0.103571 / raiva: 0.020277 surpresa: 0.064755 / tristeza: 0.482759
4	1	Olá, tenho alguma experiencia no desenvolvimento de sites, pois fiz o curso técnico em informática, então gostaria de melhorar e aprender coisas novas.	alegria alegria: 0.431392 / desgosto: 0.104940 medo: 0.023412 / raiva: 0.071261 surpresa: 0.227579 / tristeza: 0.141416
5	1	Bom dia colegas, tudo bem com vocês? Já fiz sites, mas apenas com ferramentas como o WIX, então quero aprender como funciona de verdade a criação de páginas pra Web.	tristeza alegria: 0.060954 / desgosto: 0.004633 medo: 0.004007 / raiva: 0.022501 surpresa: 0.004226 / tristeza: 0.903679
1	3	Achei bem difícil fazer o site, não consegui na primeira tentativa, mas aos poucos fui fazendo.	alegria alegria: 0.276075 / desgosto: 0.115748 medo: 0.174988 / raiva: 0.222804 surpresa: 0.025429 / tristeza: 0.184955
1	3	Sim eu também, não consegui tudo que precisava com este curso, para terminar a tarefa tive que assistir aulas no Youtube.	raiva alegria: 0.038674 / desgosto: 0.051483 medo: 0.070128 / raiva: 0.708679 surpresa: 0.025992 / tristeza: 0.105044
2	3	Achei fácil de fazer, mas não gostei muito do resultado.	tristeza alegria: 0.188811 / desgosto: 0.029723 medo: 0.214864 / raiva: 0.017596 surpresa: 0.006097 / tristeza: 0.542909
3	3	Eu achei a tarefa bem difícil, demorei até conseguir terminar. Minha principal dificuldade foi inserir figuras e deixar onde eu queria.	raiva alegria: 0.089446 / desgosto: 0.037502 medo: 0.172122 / raiva: 0.523910 surpresa: 0.075936 / tristeza: 0.101084
3	3	Também, achei o curso não explicou tudo que era preciso.	raiva alegria: 0.126045 / desgosto: 0.178579 medo: 0.190967 / raiva: 0.448544 surpresa: 0.016451 / tristeza: 0.039413
3	3	Eu precisaria de mais aulas, para melhorar.	medo alegria: 0.298963 / desgosto: 0.067842 medo: 0.322595 / raiva: 0.191737 surpresa: 0.066436 / tristeza: 0.052427
3	3	Fiquei com muitas dúvidas quanto as tabelas e figuras.	tristeza alegria: 0.076945 / desgosto: 0.065843 medo: 0.110539 / raiva: 0.065700 surpresa: 0.209820 / tristeza: 0.471152
4	3	A tarefa foi simples, não tive dificuldades em fazer, e a pagina ficou bem legal.	surpresa alegria: 0.274905 / desgosto: 0.055009 medo: 0.130879 / raiva: 0.166642 surpresa: 0.295703 / tristeza: 0.076863

5	3	Bom dia, achei a tarefa fácil e aprendi como fazer a página desde o início, então gostei foi boa a atividade.	tristeza alegria: 0.010421 / desgosto: 0.039144 medo: 0.020893 / raiva: 0.080763 surpresa: 0.027984 / tristeza: 0.820795
5	3	Não tive as mesmas dificuldades, mas também pesquisei na internet algumas dúvidas.	surpresa alegria: 0.118524 / desgosto: 0.110139 medo: 0.095486 / raiva: 0.290642 surpresa: 0.305738 / tristeza: 0.079471
5	3	Gostei do resultado ficou bem bonita!	alegria alegria: 0.481508 / desgosto: 0.031832 medo: 0.075734 / raiva: 0.031762 surpresa: 0.056362 / tristeza: 0.322801

*Os erros de português e palavras abreviadas foram corrigidos para que as postagens ficassem mais legíveis
Fonte: Autor

Para uma melhor compreensão das informações contidas no Quadro 3 foi elaborado um gráfico (Figura 2) que apresenta a emoção predominante nas postagens de cada um dos alunos, para isso foi aplicada uma média simples nos resultados apresentados pelo minerador para as publicações dos estudantes, dessa forma é possível visualizar de forma geral como estes estavam se sentindo. Por exemplo, o aluno 4 realizou duas postagens, então foram somados os valores obtidos pelo minerador a cada emoção e depois dividido por dois.

Figura 2: Média do Resultado do Minerador por Aluno



Fonte: Autor

Como pode ser percebido, no Quadro 3, o algoritmo, além de indicar a emoção predominante, também indica a probabilidade de o texto pertencer a outras categorias. Para um melhor entendimento, considere a primeira postagem do aluno 1 e sua classificação: apesar de surpresa ter uma probabilidade alta, por

volta de 20%, alegria teve uma porcentagem acima de 50%, mais precisamente 62%, o que traz certa segurança na classificação. Todavia, na terceira postagem do aluno 5 (penúltima linha da tabela), que foi classificada como surpresa, é difícil afirmar que seja realmente essa a emoção expressada, pois surpresa apresenta 30% de probabilidade, enquanto raiva aparece com 29%. Nesse caso, a identificação tem que levar em conta as duas categorias que tiveram as maiores porcentagens. Depois de feito o processamento das postagens e a determinação das emoções predominantes em cada uma delas, foi realizada uma análise da conclusão dos alunos, na qual foi possível constatar que os alunos 1, 4 e 5 concluíram o MOOC e os alunos 2 e 3 não o concluíram. A análise sintetizada pode ser observada no Quadro 4.

Quadro 4: Síntese das Análises

Aluno	Postagem 1	Postagem 2	Postagem 3	Postagem 4	Postagem 5	STATUS
1	Alegria	Alegria	Raiva			Concluiu
2	Tristeza	Tristeza				Abandonou
3	Tristeza	Raiva	Raiva	Medo	Tristeza	Abandonou
4	Alegria	Surpresa				Concluiu
5	Tristeza	Tristeza	Surpresa	Alegria		Concluiu

Fonte: Autor

Como pode ser observado no Quadro 4, há um indicativo de que a preponderância da tristeza e da raiva ocasionam o abandono, e que a surpresa e a alegria indicam a conclusão. Contudo, as demais emoções com probabilidades altas, na classificação, também devem ser avaliadas, e mais alunos investigados. À vista disso, mais estudos precisam ser desenvolvidos para que esses resultados possam ser generalizados, dado tratar-se de estudo preliminar, que apresenta alguns indícios com potencial para serem validados e analisados com mais profundidade, em especial considerando um número maior de postagens, principal elemento para que seja possível a generalização dos resultados apresentados nesta investigação. Todavia, apesar de os resultados ainda serem superficiais, é cabível dizer que o minerador de emoções pode constituir uma ferramenta interessante para que os professores/tutores tenham a possibilidade de identificar alunos que estejam desmotivados antes de desistirem e possam planejar ações para que esses concluam o MOOC.

6. Conclusões

O presente trabalho teve como objetivo principal a implementação de um minerador de emoções para postagens de fóruns de discussão de MOOCs, e, para validar sua aplicabilidade, realizou-se um estudo preliminar com as postagens de cinco alunos de um curso MOOC de uma plataforma brasileira. O estudo buscou também identificar quais sentimentos positivos e/ou negativos extraídos das postagens dos alunos refletem na conclusão ou abandono dos cursos.

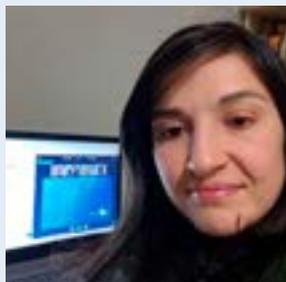
Foi possível, com a condução dessa pesquisa, identificar que o minerador apresenta uma acurácia aceitável na classificação das postagens dos alunos, em torno de 40%, para as 6 emoções especificadas. Também foi possível identificar que, se as postagens dos alunos forem categorizadas com emoções de tristeza e raiva, essas são indícios de que o aluno pode abandonar o curso, e, se forem classificadas com emoções de alegria ou surpresa, essas são indicativas de que o aluno irá terminar o curso.

Como trabalhos futuros, pretende-se implementar uma interface para que professores/tutores de MOOCs possam carregar todas as postagens de um fórum e o minerador processe a emoção predominante de cada aluno no decorrer do fórum. Além disso, espera-se realizar estudos com mais alunos, para que seja possível validar os resultados preliminares identificados nesta pesquisa.

Agradecimentos

O presente trabalho foi realizado com apoio do Instituto Federal de Educação, Ciência e Tecnologia do Rio Grande do Sul.

Biodados



SOUZA, V. F. é doutoranda no Programa de Pós-Graduação em Informática na Educação (PPGIE) da Universidade Federal do Rio Grande do Sul (UFRGS). Mestre em Informática pelo Programa de Pós-Graduação em Informática (PPGI) da Universidade Tecnológica Federal do Paraná (UTFPR), na área de Computação Aplicada, e com ênfase em Engenharia de Software. Possui Especialização em Educação Especial Inclusiva, com ênfase em Tecnologias Assistivas pela Universidade Estadual do Norte do Paraná (UENP). É graduada em Sistemas de Informação também pela UENP - Bacharel em Sistemas de Informação e Licenciada em Computação. Também Completou a Licenciatura em Matemática no Programa de Formação Pedagógica, pela UTFPR. Atualmente é docente dedicação exclusiva no Instituto Federal de Educação, Ciência e Tecnologia do Rio Grande do Sul (IFRS) no Campus Ibirubá, e está em afastamento para a realização do Doutorado. Ministra aulas nos Cursos de Ciência da Computação, Técnico em Informática Integrado do Ensino Médio, Licenciatura em Matemática e Especialização em Ensino de Linguagens e suas Tecnologias.

É membro do Grupo de Pesquisa Computação Interdisciplinar Alto Jacuí. Tem interesse nas áreas de Mineração de Dados Educacionais, Learning Analytics, Inteligência Artificial - Machine Learning e Deep Learning, Sistemas Digitais e Robótica Educacional.

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

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

Referências

DICIONÁRIO ONLINE. 2020. Disponível em: <https://www.dicio.com.br/>. Acesso em: 19 abr. 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 Cuadernos 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.