



ENSINO DURANTE A PANDEMIA: ANÁLISE DOS SENTIMENTOS DOS ESTUDANTES ATRAVÉS DO PROCESSAMENTO DE LINGUAGEM NATURAL

DOI: 10.37702/2175-957X.COBENGE.2022.4076

JOSE TARCISIO FRANCO DE CAMARGO - jtfc@bol.com.br
Centro Regional Universitário de Espírito Santo do Pinhal

Gilmar Barreto - gbarreto@dsif.fee.unicamp.br
Universidade Estadual de Campinas

ELIANA ANUNCIATO FRANCO DE CAMARGO - eafcamargo@yahoo.com.br
Centro Regional Universitário de Espírito Santo do Pinhal

Estéfano Vizconde Veraszto - estefanovv@gmail.com
UFSCar

JORGE CANDIDO - j_candido@uol.com.br
UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ

Resumo: *A pandemia de Covid-19 nos colocou em um momento histórico, singular para a educação no Brasil, onde somos levados a refletir sobre as atuais estratégias de ensino-aprendizagem. Por sua vez, para que esta reflexão apresente resultados sólidos, é necessário conhecer as opiniões e sentimentos dos alunos, que constituem o foco da educação em todos os seus níveis. Contudo, em um período de distanciamento social e oferta de ensino online, identificar sentimentos e opiniões de estudantes torna-se uma tarefa relativamente complexa. Dessa forma, uma estratégia para contornar as limitações impostas pelo distanciamento e pelo ensino online consiste na análise dos sentimentos dos discentes expressos através das redes sociais digitais. Assim, entender os sentimentos embutidos nas mídias digitais pode nortear ações em diversas áreas, como por exemplo políticas públicas de educação. Dentro deste escopo, este texto descreve um estudo a respeito dos sentimentos expressos na rede social "Twitter", relativos à educação e ao ensino durante a pandemia de Covid-19. Especificamente, esta pesquisa apresenta um modelo digital capaz de identificar as emoções básicas apresentadas em tweets que contenham as palavras "educação", "ensino" e "engenharia" no contexto da atual pandemia de Covid-19, vislumbrando o desenvolvimento de ações públicas na área da educação, em particular na educação em engenharia.*



50^o COBENGE
2022

L Congresso Brasileiro de Educação em Engenharia
e V Simpósio Internacional de Educação em Engenharia da ABENGE



"A Arte da Formação em Engenharia para
os desafios do Desenvolvimento Sustentável"

26 a 29 de SETEMBRO

Evento Online



Palavras-chave: Emoções básicas, Mídia Digital, Aprendizado Remoto.

Organização:



ENSINO DURANTE A PANDEMIA: ANÁLISE DOS SENTIMENTOS DOS ESTUDANTES ATRAVÉS DO PROCESSAMENTO DE LINGUAGEM NATURAL

1 INTRODUÇÃO

A pandemia decorrente da Covid-19 inseriu o planeta em uma crise sanitária de proporção semelhante à observada a aproximadamente cem anos atrás. Contudo, face às realidades distintas vividas no início do Século XX e, agora, no Século XXI, não é possível estudar os impactos do que ocorre atualmente através dos mesmos instrumentos tecnológicos de estudo da realidade centenária.

No presente, os meios de comunicação tornaram o acesso à informação mais rápido, bem como permitiram a produção de conteúdo em grande volume que, não necessariamente, representam a realidade. Essa dualidade, por sua vez, induz o surgimento de tensões sociais que tendem a comprometer a tomada de decisões importantes, tanto no nível pessoal quanto social, durante o período da pandemia. Tais conflitos podem ser observados em ambientes digitais, os quais são amplamente acessíveis aos envolvidos com a educação e o ensino.

Dessa forma, os conteúdos produzidos nas mídias digitais, em particular nas redes sociais, acabam por revelar sentimentos associados a temas específicos que refletem o cotidiano de alunos e professores. Assim, reconhecer e interpretar aquilo que se expressa através desses meios pode servir como mecanismo para alavancar o conhecimento a respeito da educação, bem como identificar questões que possam estar implícitas nos sentimentos expressos.

A concretização de propostas nesse sentido, porém, esbarra justamente na grande quantidade de informações que precisam ser analisadas. Além disso, questões críticas podem necessitar de análises em um intervalo de tempo muito curto, o que torna o problema ainda mais complexo.

Como exemplo, pode-se considerar o caso da educação em engenharia durante a pandemia e suas implicações sociais. Uma rápida observação das redes digitais, onde um número considerável de pessoas expressa seus sentimentos e opiniões, aponta para uma enorme quantidade de dados a respeito desta doença e suas implicações na educação e no ensino. Tais dados embutem sentimentos que, por sua vez, representam as experiências individuais em relação à pandemia. Assim, ser capaz de coletar e processar um grande volume de informações em um intervalo de tempo relativamente curto pode se tornar um diferencial para o tratamento de diversas questões que atingem a sociedade, entre elas a educação em um período pós Covid-19.

Neste contexto, este trabalho apresenta os resultados obtidos através de um estudo realizado a partir da rede social "Twitter". Especificamente, foram coletados tweets no decorrer dos meses de abril e maio de 2021, que faziam menção às palavras "educação", "ensino" e "engenharia". Tais tweets foram processados através de uma técnica computacional denominada "processamento de linguagem natural" ("*natural language processing*"), de forma a serem extraídos os sentimentos embutidos nos mesmos. Esses sentimentos foram então qualificados e quantificados de modo a fornecer um panorama geral sobre a percepção da educação e do ensino durante a pandemia através da rede social abordada.

Com a estratégia desenvolvida, espera-se contribuir para a realização de estudos mais profundos das mídias sociais, capazes de apontar com precisão suficiente os sentimentos envolvidos de modo a tratar questões de natureza diversa, que aflijam a sociedade como um todo, como por exemplo a educação e outras áreas sensíveis ao tema.

2 REFERENCIAL TEÓRICO

Através do tempo, o ser humano tem expresso sentimentos de diversas formas. Atualmente, as redes sociais digitais constituem um recurso bastante acessível para a expressão e compartilhamento de emoções. Por sua vez, a interpretação dos sentimentos embutidos nestes ambientes constitui uma rica fonte de pesquisa, a qual pode orientar ações públicas, inclusive no âmbito da educação.

Neste sentido, até meados do século XX, uma das correntes do estudo da Psicologia considerava que uma escala que varia desde "agradável" até "desagradável" era suficiente para qualificar as diferenças entre os sentimentos (EKMAN, 1957). Já ao final do século XX, essa perspectiva, que trata as emoções como fundamentalmente iguais, variando apenas quanto à intensidade de sua polarização, deu lugar ao conceito de "emoções básicas" ou sentimentos básicos (EKMAN, 1999, p.138). De acordo com esse referencial, diferentes emoções surgem a partir das interações fundamentais que ocorrem no cotidiano do ser humano, sendo universais e, portanto, independentes de culturas específicas. Nesse sentido, emoções negativas ou desagradáveis podem desencadear sentimentos como medo, raiva, repulsa ou tristeza, dentre outros. Por sua vez, emoções positivas ou agradáveis podem induzir sentimentos de alegria, alívio ou satisfação (EKMAN, 1999). Assim, as emoções básicas são organizadas em seis categorias, a saber: alegria, medo, raiva, repulsa (ou nojo), surpresa e tristeza (EKMAN, 1992a; EKMAN, 1992b). Estes sentimentos são abordados neste trabalho, tendo em vista que esta classificação continua atual (AN, 2017; CELEGHIN, 2017; MUNSHI, 2020).

Em relação à dinâmica das mídias sociais digitais, considerando-se o grande volume de informações geradas atualmente por parte significativa da população, a análise de tais sentimentos encontra um ambiente fértil com amplas possibilidades de estudo, permitindo a identificação de tendências econômicas, problemas sociais, questões educativas e de saúde pública, entre outras situações que envolvem o cotidiano das pessoas. De fato, este paradigma pode ser abordado em pesquisas que interpretam, através de técnicas de inteligência artificial, as emoções encontradas nessas redes.

Assim, este trabalho busca identificar e explorar sentimentos presentes em redes sociais a respeito da educação durante a pandemia. Por sua vez, a abordagem aqui apresentada inova ao utilizar como fundamento o modelo de predição de Naive-Bayes (RISH, 2001), o qual caracteriza-se por ser uma ferramenta matemática relativamente simples, porém eficiente para a realização de análises de grandes volumes de dados.

3 METODOLOGIA

Esta pesquisa teve início a partir da coleta de frases, de sites quaisquer da internet, que pudessem ser classificadas quanto à expressão dos seguintes sentimentos: alegria, medo, repulsa (ou nojo), raiva, surpresa e tristeza. Ao todo foram coletadas 5400 frases que, em sequência, foram rotuladas diretamente pelos pesquisadores envolvidos neste trabalho de acordo com os sentimentos indicados previamente. Esta classificação deu



origem a uma base de conhecimento na qual as frases encontraram-se distribuídas de acordo com as quantidades indicadas no Quadro 1.

Quadro 1: Distribuição das frases da base por sentimento.

Sentimento	Quantidade
Alegria	990
Medo	882
Repulsa	882
Raiva	882
Surpresa	882
Tristeza	882

Fonte: Elaborado pelos autores.

Tendo em vista que os sentimentos apresentados no Quadro 1 possuem um caráter predominantemente negativo, optou-se por um número maior de frases rotuladas como "alegria", de forma a promover um relativo balanço entre sentimentos positivos e negativos. O Quadro 2, apresenta alguns exemplos de frases rotuladas que se encontram na base de conhecimento.

Quadro 2: Exemplos de frases rotuladas na base de conhecimento.

Frase	Sentimento
<i>"vim na escola hj o só pra comer, delícia"</i>	Alegria
<i>"tenho receio do contrário, que é demonstrar que não sou muito estudioso"</i>	Medo
<i>"só no país das bananas mesmo"</i>	Repulsa
<i>"o diploma, q é só um atestado de cretino com autorização pra mentir, nem é preciso"</i>	Raiva
<i>"caramba, não fui avisado que hoje teria prova"</i>	Surpresa
<i>"não tenho motivo para seguir em frente"</i>	Tristeza

Fonte: Elaborado pelos autores.

Após a criação deste banco de conhecimento, o mesmo foi utilizado para o treinamento de um classificador de sentimentos. O objetivo deste classificador é realizar a rotulação automática de um número relativamente grande de frases para posterior análise e interpretação de seus contextos.



Foi utilizado um classificador do tipo "Naive-Bayes", que é uma ferramenta de aprendizagem de máquina baseada em "aprendizagem supervisionada". Este tipo de aprendizagem prevê a construção de uma estrutura de inteligência artificial, a qual é treinada e avaliada previamente à sua utilização. Este classificador foi implementado através da linguagem de programação Python, com o auxílio da biblioteca "nltk" (*Natural Language Toolkit*).

Optou-se pela linguagem de programação Python para a construção do classificador pois esta é considerada de altíssimo nível, de fácil aprendizagem e utilização, dispondo de muitas bibliotecas de apoio, sendo adequada aos propósitos desta pesquisa. Por sua vez, a biblioteca "nltk" consiste em uma ferramenta computacional de inteligência artificial, voltada à aprendizagem de máquina (*machine learning*), direcionada ao processamento e análise de linguagem natural, ou seja, aplicada à linguagem utilizada de forma cotidiana por seres humanos. Um dos recursos da biblioteca "nltk" permite a análise de sentimentos embutidos na linguagem escrita, que pode ser realizada através de seu classificador de Naive-Bayes.

O processo de desenvolvimento do classificador de sentimentos tem início com a divisão da base de conhecimento em dois conjuntos, direcionados para treinamento e avaliação do classificador. Para que este processo ocorra sem vícios, as frases da base de conhecimento são distribuídas de forma aleatória entre tais conjuntos de treinamento e avaliação. Neste trabalho, o conjunto de treinamento foi composto por 80% das frases da base de conhecimento e o conjunto de avaliação foi composto por 20% das frases da base de conhecimento. A partir da construção dos conjuntos de treinamento e avaliação, o software identifica as frequências das palavras mais comuns nestes conjuntos, realizando paralelamente um processo de *stemming* nestas palavras.

O passo seguinte consiste na extração das características mais informativas (*most informative features*) das frases. A Figura 1 ilustra os resultados desta etapa.

Figura 1: Características mais informativas obtidas no processo de treinamento.
Características mais informativas

med = True	medo : raiva =	146.2 : 1.0
trist = True	triste : alegri =	97.7 : 1.0
raiv = True	raiva : triste =	66.6 : 1.0
surpr = True	surpre : raiva =	55.6 : 1.0
pasm = True	surpre : medo =	35.3 : 1.0
rece = True	medo : triste =	29.9 : 1.0
felic = True	alegri : surpr =	28.2 : 1.0
feliz = True	alegri : triste =	26.4 : 1.0
fiq = True	surpre : alegri =	23.1 : 1.0
amo = True	alegri : triste =	22.8 : 1.0

Fonte: Elaborado pelos autores.

De acordo com a Figura 1, se a raiz "med" estiver presente em uma determinada frase, então a chance desta frase ser rotulada como "medo" ao invés de "raiva" é de 146,2 para 1. Por sua vez, se a raiz "amo" estiver presente em uma frase, então a probabilidade desta frase ser rotulada como "alegria" ao invés de "tristeza" é de 22,8 para 1.

Ao combinar os resultados da etapa de treinamento com a etapa de avaliação, o software apresenta a "matriz de confusão" do modelo. Para a simulação em discussão, a Figura 2 apresenta a matriz de confusão resultante.

Figura 3: Matriz de confusão obtida para o classificador construído.

Matriz de confusão:

	s	t			
a		u	r		
l		r	i		
e		r	p	s	
g	m	n	a	r	t
r	e	o	i	e	e
i	d	j	v	s	z
a	o	o	a	a	a

alegria	<174>	3	13	3	6	7
medo	5	<145>	6	10	15	9
nojo	2	4	<113>	23	12	15
raiva	9	6	16	<121>	11	15
surpresa	14	9	14	9	<120>	3
tristeza	8	9	10	14	4	<123>

(linha = referência; coluna = teste)

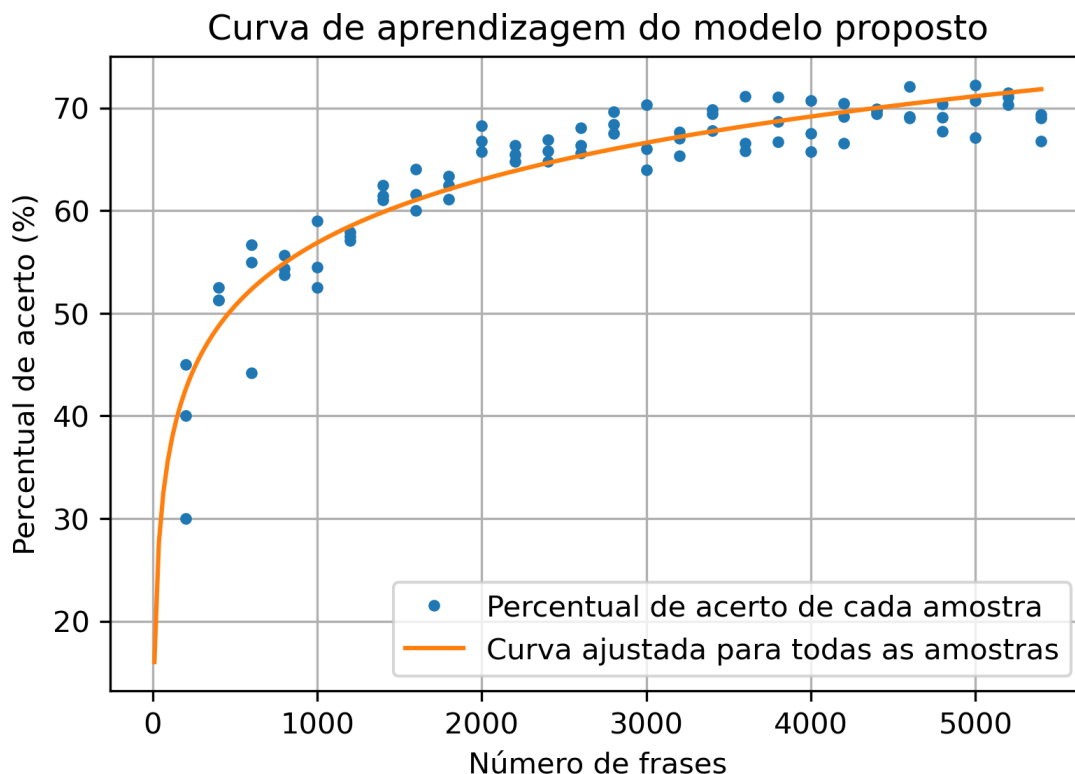
Fonte: Elaborado pelos autores.

Esta matriz permite verificar que, do conjunto de avaliação, 206 frases possuíam o rótulo "alegria" (soma das ocorrências na linha horizontal rotulada como "alegria"). Durante a fase de avaliação, 174 destas frases foram corretamente identificadas com pertencendo à classe "alegria". Por sua vez, foram equivocadamente rotuladas: 3 frases como "medo", 13 como "nojo", 3 como "raiva", 6 como "surpresa" e 7 como "tristeza". O mesmo raciocínio é válido para a análise dos resultados relativos aos demais sentimentos.

Finalmente, após a conclusão dos procedimentos de treinamento e avaliação, o algoritmo retorna o índice de acertos alcançado. Nessa simulação, o índice de acertos foi de aproximadamente 73,70%, indicando que, das 1080 frases do conjunto de avaliação, 796 foram rotuladas corretamente, conforme pode ser observado através da matriz de confusão (Figura 2).

A seleção aleatória das frases dos conjuntos de treinamento e avaliação pode levar a variações nos resultados obtidos. Além disso, o tamanho da base de conhecimento também provoca alterações na qualidade dos resultados. Para dimensionar estas oscilações e obter uma visão mais abrangente da eficiência do algoritmo, 81 simulações foram realizadas sobre a base de conhecimento. Foram efetuadas três simulações distintas para bases contendo grupos de 200, 400, 600, ... , 5000, 5200, 5400 frases. A Figura 3 apresenta a evolução da eficiência do algoritmo em função da quantidade de frases presentes na base de conhecimento.

Figura 3: Eficiência do algoritmo em função da quantidade de frases da base de conhecimento.



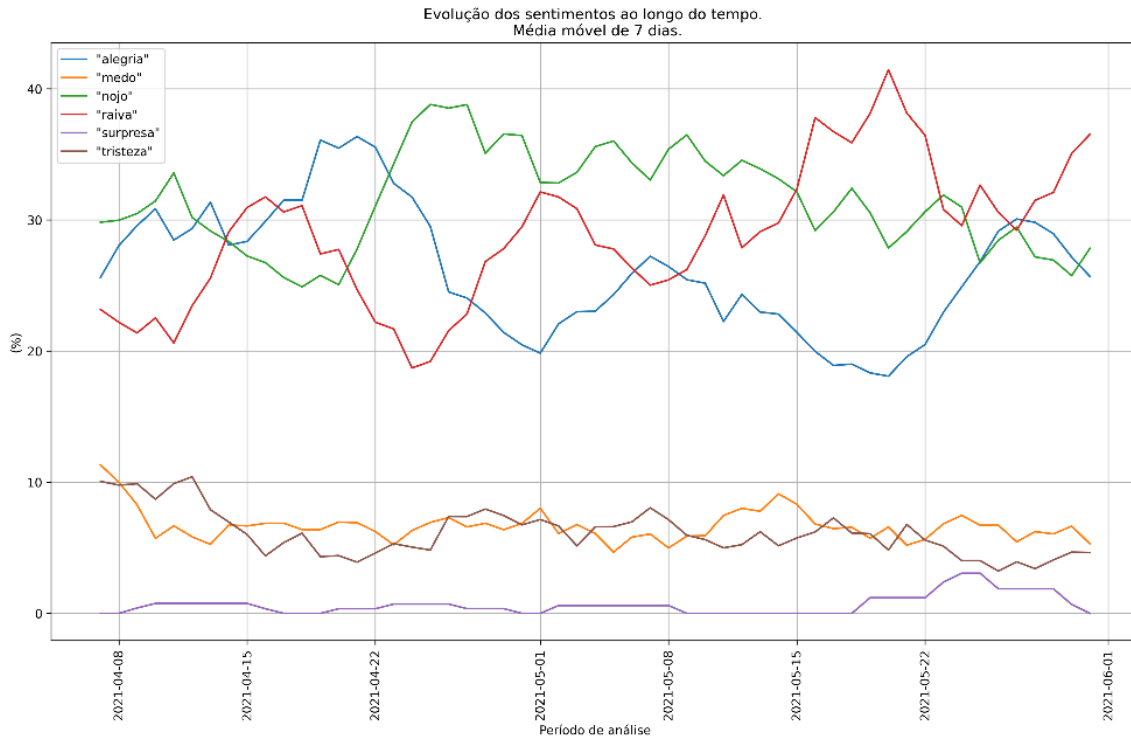
Fonte: Elaborado pelos autores.

Na Figura 3 observa-se uma curva com aspecto logarítmico que apresenta estabilização ao se aproximar de um percentual de acerto em torno de 70%. Empiricamente, nota-se esta estabilização a partir de 3400 frases. Contudo, apesar de haver uma determinada estabilidade a partir de 3400 frases presentes na base de conhecimento, observou-se também, empiricamente, que uma maior quantidade de frases na base implica em uma maior probabilidade de identificação correta do sentimento embutido em uma determinada frase.

4 RESULTADOS OBTIDOS A PARTIR DA APLICAÇÃO DO CLASSIFICADOR

Neste trabalho, a rede social “Twitter” foi pesquisada no decorrer dos meses de abril e maio de 2021. A cada dia desses meses foram pesquisados tweets que fizessem menção às palavras “educação”, “ensino” ou “engenharia” no contexto da Covid-19. Estes tweets foram, por sua vez, automaticamente rotulados pelo classificador desenvolvido quanto à expressão dos sentimentos “alegria”, “medo”, “nojo”, “raiva”, “surpresa” e “tristeza”. Diariamente foram selecionadas apenas as frases que puderam ser classificadas com 95% ou mais de probabilidade de pertencerem a uma determinada classe. A Figura 4 apresenta uma síntese dos dados obtidos e processados pelo classificador.

Figura 4: Média móvel de 7 dias para a variação quantitativa dos sentimentos observados.



Fonte: Elaborado pelos autores.

A síntese indicada através da Figura 4 mostra uma visão quantitativa a respeito dos sentimentos pesquisados. Observa-se através desta figura que, no período estudado, predominam os sentimentos de "nojo", "raiva" e "alegria", em detrimento dos sentimentos de "medo", "tristeza" e "surpresa". Contudo, tais informações não apontam para a percepção qualitativa desses sentimentos. Tal percepção pode ser observada através do Quadro 3, que apresenta um recorte dos tweets coletados no período de estudo e apresenta uma visão microscópica dos sentimentos.

Análises macroscópica e microscópica dos resultados mostram um forte sentimento negativo em relação à associação existente entre a Covid-19 e o ensino/educação durante a pandemia. Nesse contexto, um estudo particular em relação aos tweets classificados como pertencentes à classe "Alegria", de fato demonstram na maioria das vezes expressões de sarcasmo e ironia.

Quadro 3: Recorte dos tweets rotulados pelo classificador.

Data	Frase/segmento do tweet	Rótulo
01/04/2021	<i>"feliz é o jovem que terminou o ensino médio antes da pandemia"</i>	Alegria
10/04/2021	<i>"juro q comecei meu ensino médio exatamente com esse pensamento, mas ai veio pandemia e os krl, entrei em pânico pq não tava conseguindo aprender nem o básico por ead e aqui estou eu em pânico ainda"</i>	Medo
20/04/2021	<i>"os governistas e os falsos defensores da educação querem reabrir escolas, impedir o direito à greve, impondo o ensino como atividade essencial presencial, no auge da pandemia"</i>	Nojo
30/04/2021	<i>"a reclamação não é qto à qualidade do ensino, dos empresários que ganharam muito e pagaram pouco para professores demitidos durante a pandemia, o problema para esse energúmeno é tão somente: pobre ter acesso a educação"</i>	Raiva
18/05/2021	<i>"minha mãe eh prof q ta tendo ensino hibrido na minha escola, ai ontem ela falou com uma prof e ontem a msm prof fez um teste pra covid e deu positivo "</i>	Surpresa
27/05/2021	<i>"rui costa lamenta morte de diretor do núcleo de educação em saj: mais um pai de família vítima da covid"</i>	Tristeza

Fonte: Elaborado pelos autores.

5 CONSIDERAÇÕES GERAIS

O estudo tratado neste artigo apresentou a aplicação de uma ferramenta de inteligência artificial na identificação de sentimentos observados na linguagem escrita. Especificamente, o uso da ferramenta de processamento de linguagem natural mostrou-se capaz de reconhecer sentimentos presentes em textos de uma rede social. Através da coleta de *tweets* publicados no decorrer dos meses de abril de 2021 e maio de 2021, o classificador de sentimentos desenvolvido foi capaz de identificar, com precisão igual ou

superior a 95%, frases que expressam sentimentos de alegria, medo, nojo, raiva, surpresa e tristeza. O instrumento de classificação desenvolvido também demonstrou ser eficiente para o tratamento de um grande volume de informações em intervalos de tempo curtos, adequando-se perfeitamente à exploração de sentimentos presentes em redes sociais digitais. Adicionalmente, esta ferramenta é bastante flexível, podendo explorar outros contextos além de sentimentos presentes na linguagem escrita nos meios digitais.

Referências bibliográficas

AN, S.; JI, L.; MARKS, M.; ZHANG, Z. Two Sides of Emotion: Exploring Positivity and Negativity in Six Basic Emotions across Cultures. **Frontiers in Psychology**, v. 8, n. 610, 2017. Disponível em: <https://www.frontiersin.org/articles/10.3389/fpsyg.2017.00610/full> Acessado em: 04 dez. 2020.

CELEGHIN, A.; DIANO, M.; BAGNIS, A.; VIOLA, M.; TAMIETTO, M. Basic Emotions in Human Neuroscience: Neuroimaging and Beyond. **Frontiers in Psychology**, v. 8, n. 1432, 2017. Disponível em: <https://www.frontiersin.org/articles/10.3389/fpsyg.2017.01432/full> Acessado em: 04 dez. 2020.

EKMAN, P. A Methodological discussion of nonverbal behavior. **Journal of Psychology**, v. 43, p. 141-149, 1957.

EKMAN, P. An argument for basic emotions. **Cognition and Emotion**, v. 6, n. 3-4, p. 169-200, 1992.

EKMAN, P. Are there basic emotions? **Psychological Review**, v. 99, n. 3, p. 550–553, 1992.

EKMAN, P. Basic emotions. In: DALGLEISH, T.; POWER, M. (ed.) **Handbook of cognition and emotion**. Londres: John Wiley and Sons Ltd., 1999.

MUNSHI A. *et al.* Modeling the Relationships Between Basic and Achievement Emotions in Computer-Based Learning Environments. In: BITTENCOURT I., CUKUROVA M., MULDER K., LUCKIN R., MILLÁN E. (ed.) **Artificial Intelligence in Education**. AIED 2020. Lecture Notes in Computer Science, v. 12163. Springer, Cham, 2020.

RISH, I. An empirical study of the naive Bayes classifier. In: **Proceedings of the IJCAI 2001 – International Joint Conferences in Artificial Intelligence Organization: Workshop on empirical methods in artificial intelligence**, v. 3, n. 22, p. 41-46, 2001.

TEACHING DURING THE PANDEMIC: ANALYSIS OF STUDENTS' FEELINGS THROUGH NATURAL LANGUAGE PROCESSING

Abstract: *Digital social networks are a rich source of research on events that build people's daily lives. Such events express feelings that often have a universal context and, therefore, are common to the most varied cultures. Thus, understanding the feelings embedded in digital media as well as people's perception of certain subjects can guide actions in several areas, such as public education policies. So, this article describes a study about the feelings and perceptions expressed on the social network Twitter related to the students' perception concerning the pandemic and remote learning. Specifically, this research presents a digital model capable of identifying the basic emotions presented in tweets that contain the words "education", "teaching" or "engineering" in the context of the Covid-19 pandemic, envisioning the development of public actions in the area of education from the results obtained. To be able to analyze the feelings and perceptions presented in a relatively large volume of data obtained in a short time interval, this research used the Naive-Bayes prediction method, which presents relative simplicity and has shown good efficiency in real-time rankings. The results demonstrate the feasibility of using this model as well as consistently express the feelings related to engineering education during the pandemic, in Portuguese, disseminated on Twitter. It is concluded that the proposed model is fully viable for the study of large volumes of information in real-time, flexible to perform different types of classification, providing good reliability to the synthesized results.*

Keywords: *basic emotions, digital media, remote learning.*