

Utilização de Recursos Linguísticos para Classificação Automática de Mensagens de Feedback

Anderson Pinheiro Cavalcanti^{1,4}, Rafael Ferreira Mello^{2,3}, Péricles Miranda², André Nascimento², Fred Freitas¹

¹Centro de Informática – Universidade Federal de Pernambuco (UFPE)
Recife – PE – Brasil

²Departamento de Computação – Universidade Federal Rural de Pernambuco (UFRPE)
Recife – PE – Brasil

³Cesar School – Recife – PE – Brasil

⁴SiDi – Recife – PE – Brasil

{apc, fred}@cin.ufpe.br

{rafael.mello, pericles.miranda, andre.camara}@ufrpe.br

Abstract. *Feedback plays a significant role in a student's learning process. It allows students to identify weaknesses and improve self-regulation. However, studies show that this is an area of great dissatisfaction in higher education. With the increasing number of course participants, providing effective feedback is becoming an increasingly challenging task. Therefore, this article explores the use of automated content analysis to examine instructor-provided feedback based on well-regarded models from the literature that provide good practices and rank feedback at different levels. For this, binary classifiers were trained using the XGBoost algorithm together with linguistic resources that extract characteristics from the feedback messages. The results indicate effective classification performance for both good practices and feedback levels, reaching accuracy values of up to 0.89 in the test set.*

Resumo. *O feedback possui um papel significativo no processo de aprendizagem de um aluno. Permite que os alunos identifiquem os pontos fracos e melhorem a autorregulação. No entanto, estudos mostram que esta é uma área de grande insatisfação no ensino superior. Com o número cada vez maior de participantes em cursos, fornecer feedback eficaz está se tornando uma tarefa cada vez mais desafiadora. Portanto, este artigo explora o uso da análise de conteúdo automatizada para examinar o feedback fornecido por instrutores com base em modelos conceituados da literatura que fornecem boas práticas e classificam os feedbacks em níveis diferentes. Para isso, foram treinados classificadores binários usando o algoritmo XGBoost em conjunto com recursos linguísticos que extraem características das mensagens de feedback. Os resultados indicam desempenho de classificação eficaz tanto para as boas práticas quanto para os níveis de feedback atingindo valores de até 0,89 de acurácia no conjunto de teste.*

1. Introdução

Uma educação de qualidade requer atenção e apoio pessoal. Um elemento crucial para isso é o feedback, cuja definição muda dependendo da literatura de pesquisa educacio-

nal. Neste artigo, o feedback é conceituado como o fluxo de informações de um agente para outro a respeito de uma decisão do aluno [Hattie and Timperley 2007]. Vários estudos destacam o papel significativo que o feedback desempenha no processo de aprendizagem [Batista et al. 2018, Cavalcanti et al. 2021]. [Price et al. 2010] afirmam que o feedback é o componente mais crítico das avaliações. O feedback direciona os alunos para o tipo apropriado de estudo ou prática e ajuda os indivíduos a reconhecer áreas de deficiência, que podem ser usadas para aprimorar as táticas e estratégias de aprendizagem [Parikh et al. 2001, Weaver 2006]. [Dweck 1999] relata que o feedback pode afetar a motivação do aluno, bem como os conteúdos e habilidades aprendidas. [Laurillard 1993] enfatizou que a ação sem feedback é completamente improdutiva para o aluno.

Apesar do reconhecimento generalizado da importância do feedback para a aprendizagem, grande parte da literatura atual indica a difusão do feedback de baixa qualidade no ensino superior [Hattie and Gan 2011]. A qualidade do feedback é consistentemente avaliada como uma das maiores causas de insatisfação para estudantes do ensino superior [Ferguson 2011]. [Weaver 2006] relata que, embora os acadêmicos reconheçam o valor do feedback para facilitar a aprendizagem, eles consideram os comentários dos instrutores incompreensíveis e ineficazes. [Ferguson 2011] identifica a falta do fornecimento de feedback oportuno, expectativas pouco claras e pouca utilidade como as principais preocupações entre os alunos.

À medida que as instituições de ensino superior adotam a tecnologia, há um portfólio crescente de abordagens que utilizam a coleta de dados para melhorar os processos de aprendizagem. De acordo com a revisão sistemática realizada por [Cavalcanti et al. 2021] vários trabalhos estão explorando ativamente soluções de feedback automatizado que podem permitir que os instrutores identifiquem e empreguem boas práticas de feedback de maneira eficiente e aumentem a velocidade de entrega de feedback aos alunos. Nesse sentido, alguns estudos examinaram o uso de métodos de mineração de dados para gerar feedback textual automatizado [Liu et al. 2017, Ma et al. 2017, Villalón et al. 2008]. Essas análises são frequentemente limitadas a áreas específicas de domínio, como programação de computadores ou redação, ou falta de base na teoria educacional.

Na literatura é possível encontrar trabalhos que apresentam abordagens para analisar feedback automaticamente [Osakwe et al. 2021, Cavalcanti et al. 2020b]. Contudo, são abordagens que focam em textos de feedback da língua inglesa ou utilizam poucos recursos linguísticos para extrair informações dos textos.

Diante deste contexto, esse artigo tem como objetivo propor uma abordagem que combina diferentes recursos linguísticos para classificar mensagens de feedback usando o algoritmo XGBoost com base em conceituados modelos de feedback da literatura [Hattie and Timperley 2007, Nicol and Macfarlane-Dick 2006]. Os resultados indicam um desempenho de classificação eficaz para ambos os modelos de feedback alcançando melhor desempenho do que outros trabalhos da literatura.

2. Trabalhos Relacionados

Alguns trabalhos teorizam que o objetivo principal do feedback é reduzir a discrepância entre o conhecimento/desempenho do exercício e uma meta [Hattie and Timperley 2007]. Portanto, o feedback é frequentemente examinado através da autorregulação da apren-

dizagem [Clark 2012, Nicol and Macfarlane-Dick 2006]. [Butler and Winne 1995] reconhecem que os alunos autorregulados são os alunos mais eficazes. A autorregulação é um ciclo de estabelecimento de metas de construção de conhecimento, seleção de estratégias que maximizem o progresso em direção a tais metas e monitoramento do progresso com a possibilidade de alterar estratégias dependendo do nível de progressão [Clark 2012].

De acordo com [Nicol and Macfarlane-Dick 2006], a boa prática de feedback é amplamente definida como qualquer estratégia ou conteúdo que possa aumentar a capacidade dos alunos de autorregular seu desempenho de aprendizagem. [Nicol and Macfarlane-Dick 2006] propôs sete princípios gerais de boas práticas de feedback para auxiliar o corpo docente: (i) Ajuda a esclarecer o que é o bom desempenho (objetivos, critérios, padrões esperados), (ii) Facilita o desenvolvimento de autoavaliação (reflexão) na aprendizagem, (iii) Oferece informações de alta qualidade aos alunos sobre sua aprendizagem, (iv) Incentiva o professor e o diálogo com os colegas sobre a aprendizagem, (v) Incentiva crenças motivacionais positivas e auto-estima, (vi) Oferece oportunidades para fechar a lacuna entre o desempenho desejado, (vii) Fornece informações aos professores que podem ser usadas para ajudar a moldar o ensino.

[Hattie and Timperley 2007] analisaram várias condições que poderiam maximizar os efeitos positivos do feedback na aprendizagem, incluindo o aumento da consciência do aluno sobre uma meta geral de aprendizagem, o progresso em direção à meta e as metas subsequentes necessárias para atingir o objetivo principal. Assim, [Hattie and Timperley 2007] propuseram quatro perspectivas, propostas como níveis, que o feedback deve abordar a fim de melhorar sua eficácia. Eles postularam que seu modelo é mais adequado para examinar o feedback textual porque o modelo é focado em aspectos relacionados a tarefas de aprendizagem, processo de aprendizagem e autorregulação do aluno. A Tabela 1 mostra os níveis propostos por [Hattie and Timperley 2007].

Tabela 1. Níveis de feedback [Hattie and Timperley 2007]

#	Nível	Descrição
FT	Feedback sobre a tarefa	O feedback pode ser sobre uma tarefa, como se o trabalho está correto ou incorreto, pode incluir instruções para mais informações ou informações diferentes.
FP	Feedback sobre o processamento da tarefa	O feedback pode ser direcionado ao processo usado para criar um produto ou concluir uma tarefa, é mais direcionado ao processamento de informações ou processos de aprendizagem que requerem compreensão ou conclusão da tarefa.
FR	Feedback sobre autorregulação	O feedback para os alunos pode ser focado no nível de autorregulação, incluindo maior autoavaliação ou habilidades de confiança, que podem ter grandes influências na autoeficácia, na proficiência autorregulatória e nas crenças pessoais dos alunos.
FS	Feedback pessoal.	O feedback pode ser pessoal no sentido de que é direcionado a si mesmo. Frequentemente, não está relacionado ao desempenho da tarefa.

[Cavalcanti et al. 2020b] propôs uma análise de conteúdo do texto de feedback fornecido pelos instrutores com base nas boas práticas de feedback propostas por [Nicol and Macfarlane-Dick 2006]. Os autores se concentraram em analisar a qualidade do feedback extraído das avaliações coletadas em um curso online oferecido em uma

instituição de ensino superior brasileira. Um dos problemas desse trabalho é a divisão dos feedbacks em classes binárias (classe 0: não tem boa prática; classe 1: tem pelo menos 1 boa prática). Esse tipo de classificação pode não ser totalmente eficaz, pois o professor apenas tem a informação se o texto tem ou não boa prática. Para que o professor consiga fornecer um feedback de qualidade para os alunos, é necessário que ele saiba quais dentre as 7 boas práticas o seu texto está seguindo ou não [Nicol and Macfarlane-Dick 2006]. Além disso, a combinação de ambas as teorias pode ajudar o professor no fornecimento de um feedback de qualidade [Cavalcanti et al. 2020a].

Para treinar um classificador é necessário transformar o texto de feedback em características numéricas. Existem ferramentas linguísticas que auxiliam na extração de características. O trabalho de [Camelo et al. 2020] propõe uma adaptação da ferramenta Coh-matrix para a língua Portuguesa. Essa ferramenta extrai diferentes recursos linguísticos do texto, incluindo legibilidade, coerência e coesão textual. Esses recursos podem ser utilizados para análise de diferentes tipos de textos educacionais como redações, mensagens em fóruns educacionais ou até mesmo feedback educacional. [Ferreira et al. 2020] propõem uma abordagem baseada em uma combinação de recursos tradicionais de mineração de texto e contagem de palavras extraídas com o uso de estruturas linguísticas para rotulagem automática do conteúdo das mensagens em discussões online.

O trabalho de [Cavalcanti et al. 2020c] propõe um abordagem para analisar textos de feedback automaticamente utilizando o algoritmo *Random Forest*. Os autores seguiram o modelo que classifica o feedback em quatro níveis diferentes [Hattie and Timperley 2007]. Contudo, os autores utilizaram um conjunto limitado de características, o que pode reduzir a generalização do classificador para os diferentes níveis de feedback.

Dessa forma, este trabalho difere dos trabalhos mencionados em 3 aspectos principais: (1) Classificar automaticamente as boas práticas propostas por [Nicol and Macfarlane-Dick 2006]; (2) Classificar automaticamente os níveis de feedback propostos por [Hattie and Timperley 2007]; (3) Utilizar o algoritmo XGBoost em conjunto com novos recursos linguísticos para análise das mensagens de feedback [Camelo et al. 2020].

3. Método

3.1. Conjunto de Dados

O conjunto de dados usado nesse estudo consiste em comentários de feedback fornecidos por instrutores [Cavalcanti et al. 2019]. O conjunto de dados possui 1.000 exemplos de feedback escritos em português dos cursos de Biologia e Literatura. O comprimento médio do comentário foi de 30 palavras por comentário. Os exemplos de feedback foram codificados por especialistas usando instruções propostas nos dois modelos de feedback [Nicol and Macfarlane-Dick 2006, Hattie and Timperley 2007]. Com isso, cada texto de feedback poderia ser classificado da seguinte forma: classe 0, se o feedback não pertence ao nível ou boa prática; classe 1: se o feedback pertence ao nível ou boa prática. A Tabela 2 mostra a divisão do conjunto de dados para cada boa prática e nível de feedback.

Tabela 2. Divisão da base de dados para os níveis e boas práticas.

	BP1	BP2	BP3	BP4	BP5	BP6	BP7	FT	FP	FR	FS
Classe 0	875	971	692	987	766	397	1000	112	499	992	849
Classe 1	125	29	308	13	234	603	0	888	501	8	151
Total	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000	1000

3.2. Extração de características

Foram extraídas 161 características dos textos usando o LIWC (*Linguistic Inquiry and Word Count*), o Coh-Metrix proposto por [Camelo et al. 2020], e 3 características adicionais (número de entidades nomeadas, presença de saudação e presença de expressões positivas). Essas características foram escolhidas por terem apresentado bom desempenho em trabalhos de análise de textos educacionais [Barbosa et al. 2020, Cavalcanti et al. 2020c, Barbosa et al. 2021].

O LIWC é uma ferramenta lexical que analisa palavras de grupos de texto que refletem várias dimensões psicológicas [Tausczik and Pennebaker 2010]. As principais categorias fornecidas pelo LIWC incluem processos cognitivos, processos sociais, linguagem informal, preocupações pessoais, afeto, relatividade, orientação temporal, impulsos e processos perceptivos. A contribuição relativa dessas categorias no texto oferece um perfil descritivo de vários construtos psicológicos envolvidos na escrita. O dicionário LIWC tem mais de 120.000 palavras, onde cada palavra pode ser atribuída a uma ou mais categorias. Foram extraídas 64 características do LIWC.

O Coh-Metrix é um sistema de linguística computacional que mede um conjunto de características sobre o texto que são amplamente adotadas na pesquisa educacional para avaliar a qualidade das atividades escritas [Camelo et al. 2020]. Coh-Metrix pode fornecer uma visão sobre coesão, linguagem, complexidade e legibilidade de textos [McNamara et al. 2014]. Foram extraídas 94 características do Coh-Metrix em sua versão para a língua portuguesa [Camelo et al. 2020].

Além do LIWC e Coh-Metrix, foram extraídas 3 características adicionais: número de entidades nomeadas, presença de saudação e presença de expressões positivas. O número de entidades nomeadas pode fornecer informações sobre o nível de detalhe nas mensagens de feedback entregues aos alunos. A presença de saudação e expressão positiva está relacionada aos elogios ou crenças motivacionais que o professor pode fazer a um aluno para atender a boa prática 5 [Nicol and Macfarlane-Dick 2006].

3.3. Classificação

Nesse trabalho foi utilizado o algoritmo *XGBoost* devido ao seu desempenho em comparação a outros métodos e por ser um algoritmo caixa branca [Chen and Guestrin 2016]. Foi demonstrado em [Xiao et al. 2017, Pan 2018] que o *XGBoost* supera o Random Forest em várias tarefas de classificação. O algoritmo utiliza o aumento de gradiente, que envolve modelos de combinação sequencial (neste caso, árvores de decisão) que preveem os resíduos ou erros de modelos anteriores em cada iteração para melhorar a precisão geral [Chen and Guestrin 2016].

Para desenvolver e avaliar um sistema de classificação de texto, primeiro precisa-

se definir os conjuntos de dados de treinamento e teste. Neste trabalho, os dados foram divididos em 80% para o treinamento e 20% para o teste, onde foi utilizado a validação cruzada de 10 vezes no treinamento e depois selecionado o modelo que obteve a maior acurácia para ser aplicado nos dados de teste. Como o *XGBoost* é um algoritmo baseado em árvores de decisão, é possível definir o número de árvores geradas para realização do treinamento (parâmetro *n_{tree}*). Nesse trabalho foi utilizado diferentes números de árvores para cada classificador durante o treinamento e o melhor resultado foi aplicado na base de teste.

Tabela 3. Divisão em treino e teste para cada boa prática do dataset.

		Classe 0	Classe 1	Total
BP1	Treino	700 (87.25%)	100 (12.75%)	800 (80%)
	Teste	175 (87.25%)	25 (12.75%)	200 (20%)
BP3	Treino	554 (69.25%)	246 (30.75%)	800 (80%)
	Teste	138 (69.00%)	62 (31.00%)	200 (20%)
BP5	Treino	613 (76.62%)	187 (23.38%)	800 (80%)
	Teste	153 (76.50%)	47 (23.50%)	200 (20%)
BP6	Treino	318 (39.75%)	482 (60.25%)	800 (80%)
	Teste	79 (39.50%)	121 (60.50%)	200 (20%)

Como é possível perceber pela Tabela 2, algumas classes possuem um número muito pequeno de mensagens de feedback. Por exemplo, a classe 1 da boa prática 2 (BP2) possui apenas 29 mensagens. Essa quantidade é insuficiente para treinar um modelo de aprendizado de máquina [Friedman et al. 2001]. Dessa forma, as classes BP2, BP4, BP7 e FR foram removidas do experimento, formando um novo conjunto com 4 boas práticas (BP1, BP3, BP5 e BP6) e 3 níveis (FT, FP e FS) conforme mostram as Tabelas 3 e 4, respectivamente.

Para avaliar nosso modelo, usamos as medidas de acurácia e Kappa de Cohen. A acurácia pode ser calculada usando a proporção de amostras corretamente identificadas para o número total de amostras. Além disso, o Kappa mede o acordo entre avaliadores, levando em consideração a possibilidade de acordo por acaso; assim, pode ser usado para medir o nível de concordância entre mensagens codificadas automaticamente e manualmente no conjunto de dados. O Kappa é amplamente usado na literatura de análise de conteúdo e complementa o resultado fornecido pela acurácia [Neuendorf 2016, Ferreira-Mello et al. 2019].

4. Resultados

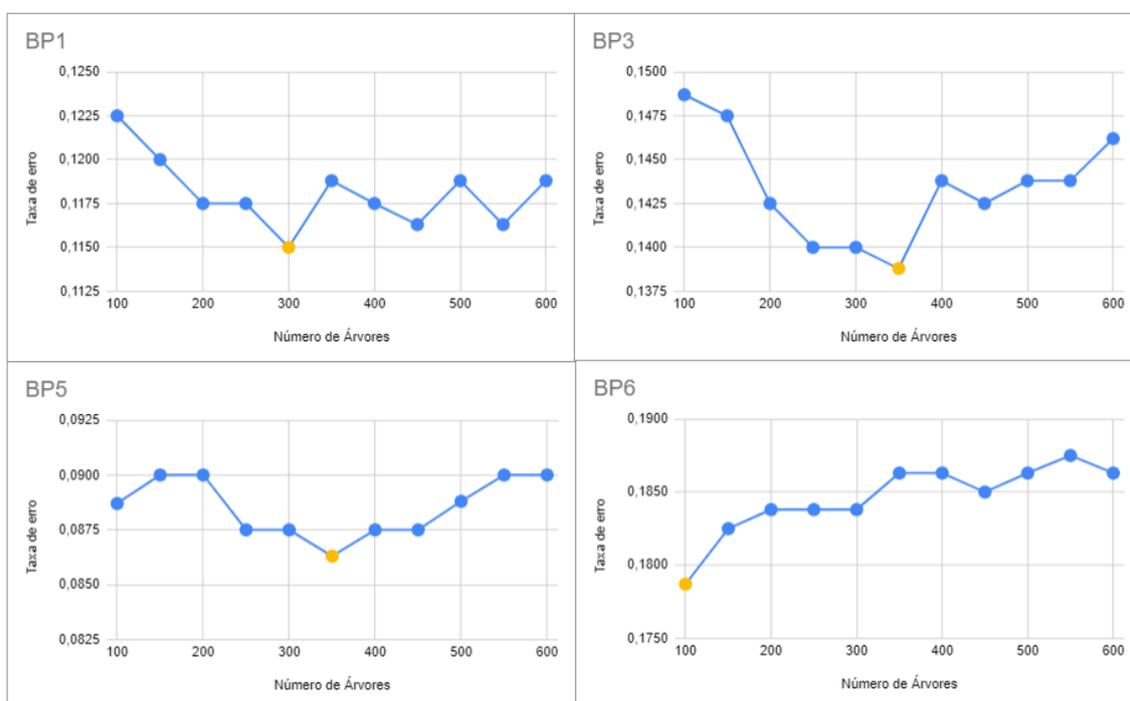
Foram analisados diferentes números de árvores (parâmetro *n_{tree}*), no intervalo entre 100 e 600, nos conjuntos de treinamento para cada boa prática e nível de feedback utilizando a técnica de validação cruzada 10 vezes. Dessa forma, o modelo com o menor erro no conjunto de treinamento foi aplicado no conjunto de teste.

4.1. Análise das boas práticas de feedback

A Figura 1 mostra os resultados obtidos no conjunto de treinamento para cada boa prática. Para a boa prática 1, o menor erro foi obtido utilizando 300 árvores com uma acurácia

Tabela 4. Divisão em treino e teste para cada nível do dataset.

		Classe 0	Classe 1	Total
FT	Treino	90 (11.25%)	710 (88.75%)	800 (80%)
	Teste	22 (11.00%)	178 (89.00%)	200 (20%)
FP	Treino	399 (49.87%)	401 (51.13%)	800 (80%)
	Teste	100 (50.00%)	100 (50.00%)	200 (20%)
FS	Treino	679 (84.87%)	121 (15.13%)	800 (80%)
	Teste	170 (85.00%)	30 (15.00%)	200 (20%)

**Figura 1. Análise do número de árvores x erro para as boas práticas.**

de 0,88 e Kappa de 0,34. Para a boa prática 3 e boa prática 5, o melhor resultado foi obtido utilizando 350 árvores com acurácia de 0,86 e 0,91 e Kappa de 0,65 e 0,75, respectivamente. Já para a boa prática 6, o melhor resultado foi obtido com 100 árvores apresentando uma acurácia de 0,82 e Kappa de 0,62.

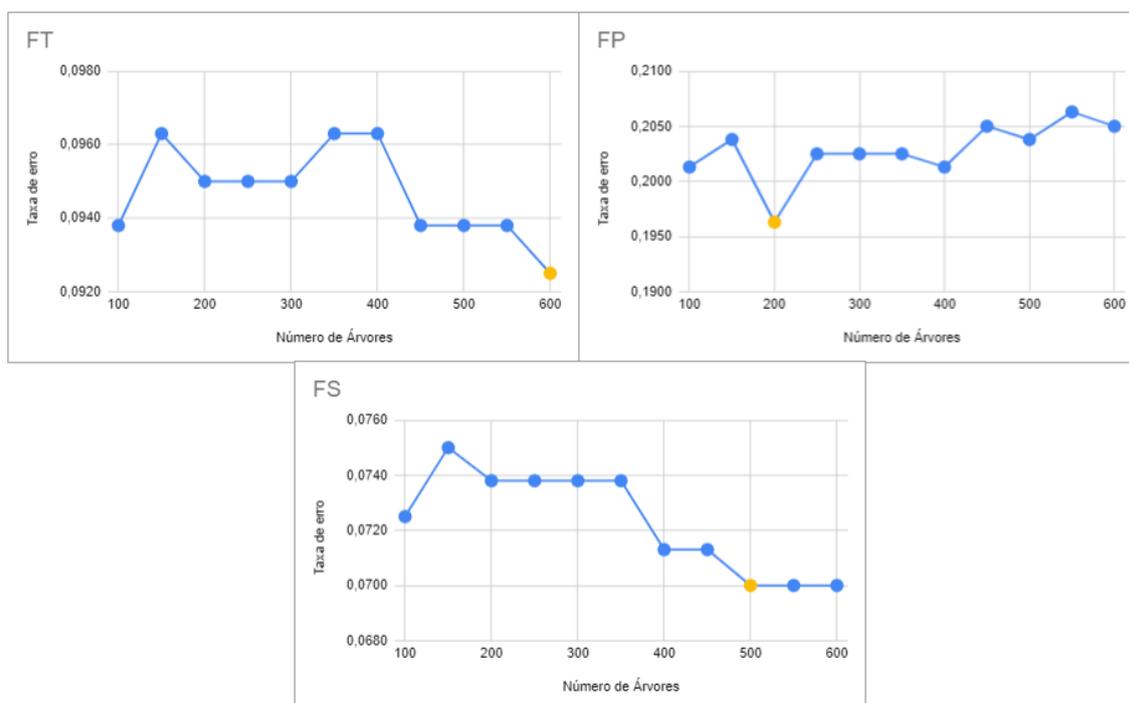
Os melhores modelos obtidos no conjunto de treinamento foram aplicados no conjunto de teste. A Tabela 5 mostra as medidas de acurácia e Kappa para cada boa prática. As boas práticas com maior acurácia foram a BP1 e a BP5, com 0,87 e 0,88, respectivamente. Entretanto, a boa prática 1 teve um valor de Kappa menor que as outras boas práticas (semelhante ao resultado obtido no conjunto de treinamento). Esse valor de Kappa pode ser atribuído ao fato de que a boa prática 1 está desbalanceada, ou seja, tem poucos exemplos da classe 1 (125 exemplos) comparando com o número de exemplos da classe 0 (875 exemplos). Dessa forma, a utilização de mais de 1 métrica de avaliação em algoritmos de aprendizagem de máquina é extremamente importante para entendermos o funcionamento do classificador [Mani and Zhang 2003].

Tabela 5. Resultados obtidos no conjunto de teste para as boas práticas.

	Acurácia	Kappa
BP1	0,87	0,26
BP3	0,84	0,60
BP5	0,88	0,66
BP6	0,81	0,60

4.2. Análise dos níveis de feedback

A Figura 2 mostra os resultados obtidos no conjunto de treinamento para cada nível de feedback. O nível de tarefa (FT) obteve o menor erro com 600 árvores, apresentando uma acurácia de 0,90 e Kappa de 0,32. Para o nível de processamento da tarefa (FP) o melhor resultado foi obtido usando 200 árvores com uma acurácia de 0,80 e Kappa de 0,60. Já para o nível pessoal (FS) foram necessárias 500 árvores para obter uma acurácia de 0,93 e Kappa de 0,70.

**Figura 2. Análise do número de árvores x erro para os níveis.**

A Tabela 6 mostra os resultados do conjunto de teste aplicando o melhor modelo obtido no conjunto treinamento. Foram obtidos os valores de 0,89 de acurácia e 0,42 de Kappa para o nível FT, 0,77 de acurácia e 0,55 de Kappa para o nível FP e 0,89 de acurácia e 0,56 de Kappa para o nível FS. A Tabela 6 também compara o nosso resultado com um estudo anterior [Cavalcanti et al. 2020c]. O trabalho de [Cavalcanti et al. 2020c] utilizou o classificador *Random Forest* em conjunto com a aplicação do algoritmo SMOTE para lidar com o desbalanceamento das classes e com um conjunto de 116 características extraídas. Nossa abordagem propõe a utilização do algoritmo XGBoost, que já lida

com o problema de desbalanceamento de classes, e um conjunto maior de 161 características extraídas. Em todos os níveis nossa abordagem superou o resultado proposto por [Cavalcanti et al. 2020c] tanto na medida de acurácia quanto na de Kappa.

Tabela 6. Resultados obtidos no conjunto de teste para os níveis de feedback.

Abordagem proposta		[Cavalcanti et al. 2020c]		
	Acurácia	Kappa	Acurácia	Kappa
FT	0,89	0,42	0,75	0,29
FP	0,77	0,55	0,64	0,28
FS	0,89	0,56	0,87	0,39

5. Conclusão

Este trabalho apresentou uma abordagem para classificação de mensagens de feedback em boas práticas e níveis propostos por [Nicol and Macfarlane-Dick 2006] e [Hattie and Timperley 2007], respectivamente. Foram extraídas 161 características das mensagens de feedback utilizando os recursos LIWC e Coh-Metrix em conjunto com 3 características adicionais. Foi utilizado o algoritmo XGBoost para classificação das mensagens em 4 boas práticas [Nicol and Macfarlane-Dick 2006] e três níveis [Hattie and Timperley 2007] diferentes. Foram analisados diferentes números de árvores (parâmetro do classificador XGBoost) no conjunto de treinamento e o classificador com menor erro foi aplicado no conjunto de testes.

Os resultados dos classificadores mostraram que a combinação dos recursos LIWC e Coh-Metrix e características adicionais foram eficazes na classificação das mensagens de feedback, atingindo acurácias acima de 0,81 nos conjuntos de teste das boas práticas e acima de 0,77 nos conjuntos de testes dos níveis. Além disso, a utilização de um número maior de características extraídas (161) em conjunto com um classificador mais recente (XGBoost) trouxe resultados significativos comparado com um trabalho relacionado. Nossa abordagem teve aumentos expressivos de 18,66% para acurácia e 44,82% para o Kappa no nível FT, 20,31% para acurácia e 96,42% para o Kappa no nível FP e 2,29% para acurácia e 43,58% para o Kappa no nível FS comparados com a abordagem de [Cavalcanti et al. 2020c]. Não foi encontrado na literatura nenhum trabalho que classificasse as boas práticas individualmente. Por esse motivo, nosso trabalho compara apenas os resultados dos níveis de feedback com o resultado obtido por [Cavalcanti et al. 2020c].

É importante salientar que os classificadores desenvolvidos nesse trabalho podem ser utilizados em conjunto com sistemas/ferramentas de ambientes virtuais de aprendizagem e dessa forma, pode auxiliar o professor na identificação automática de quais boas práticas o seu texto está seguindo e em qual nível de feedback o seu texto está classificado.

6. Limitação e trabalhos futuros

Uma das limitações deste trabalho é o desbalanceamento das classes. As boas práticas BP2, BP4 e BP7 e o nível FR não tinham exemplos suficientes na classe 1 para treinar um classificador e por isso, foram removidas dos experimentos. Além disso, o conjunto de dados é formado por mensagens de feedback de apenas 2 cursos à distância de uma

Universidade pública, e com isso, não generaliza os diferentes tipos de feedback em ambientes virtuais de aprendizagem. Como trabalhos futuros, pretendemos crescer o conjunto de dados com feedbacks de outros cursos e aplicar os classificadores em um ambiente real para auxiliar instrutores. Além disto, técnicas de balanceamento de dados (*over e under sampling*) devem ser analisadas para tentar melhorar a performance dos resultados finais.

Referências

- [Barbosa et al. 2021] Barbosa, A., Ferreira, M., Ferreira Mello, R., Dueire Lins, R., and Gasevic, D. (2021). The impact of automatic text translation on classification of online discussions for social and cognitive presences. In *LAK21: 11th International Learning Analytics and Knowledge Conference*, pages 77–87.
- [Barbosa et al. 2020] Barbosa, G., Camelo, R., Cavalcanti, A. P., Miranda, P., Mello, R. F., Kovanić, V., and Gašević, D. (2020). Towards automatic cross-language classification of cognitive presence in online discussions. In *Proceedings of the Tenth International Conference on Learning Analytics & Knowledge*, pages 605–614.
- [Batista et al. 2018] Batista, R. L., Salgado, N., and Barreto, R. (2018). Avaliação e feedback automático em educação apoiada por tecnologia: Um mapeamento sistemático da literatura. In *Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE)*, volume 29, page 178.
- [Butler and Winne 1995] Butler, D. L. and Winne, P. H. (1995). Feedback and Self-Regulated Learning: A Theoretical Synthesis. *Review of Educational Research*, 65(3):245–281.
- [Camelo et al. 2020] Camelo, R., Justino, S., and de Mello, R. F. L. (2020). Coh-matrix pt-br: uma api web de análise textual para a educação. In *Anais dos Workshops do IX Congresso Brasileiro de Informática na Educação*, pages 179–186. SBC.
- [Cavalcanti et al. 2020a] Cavalcanti, A., Rolim, V., Mello, R., and Freitas, F. (2020a). Uma análise entre boas práticas de feedback em ambientes virtuais de aprendizagem. In *Anais do XXXI Simpósio Brasileiro de Informática na Educação*, pages 962–971, Porto Alegre, RS, Brasil. SBC.
- [Cavalcanti et al. 2021] Cavalcanti, A. P., Barbosa, A., Carvalho, R., Freitas, F., Tsai, Y.-S., Gašević, D., and Mello, R. F. (2021). Automatic feedback in online learning environments: A systematic literature review. *Computers and Education: Artificial Intelligence*, 2:100027.
- [Cavalcanti et al. 2020b] Cavalcanti, A. P., de Mello, R. F. L., de Miranda, P. B. C., and de Freitas, F. L. G. (2020b). Análise automática de feedback em ambientes de aprendizagem online. In *Anais do XXXI Simpósio Brasileiro de Informática na Educação*, pages 892–901. SBC.
- [Cavalcanti et al. 2019] Cavalcanti, A. P., de Mello, R. F. L., Rolim, V., André, M., Freitas, F., and Gašević, D. (2019). An analysis of the use of good feedback practices in online learning courses. In *2019 IEEE 19th International Conference on Advanced Learning Technologies (ICALT)*, volume 2161, pages 153–157. IEEE.
- [Cavalcanti et al. 2020c] Cavalcanti, A. P., Diego, A., Mello, R. F., Mangaroska, K., Nascimento, A., Freitas, F., and Gašević, D. (2020c). How good is my feedback? a content analysis of written feedback. In *Proceedings of the 10th International Conference on Learning Analytics and Knowledge - LAK*. ACM.

- [Chen and Guestrin 2016] Chen, T. and Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '16*, pages 785–794, San Francisco, California, USA. ACM.
- [Clark 2012] Clark, I. (2012). Formative Assessment: Assessment Is for Self-regulated Learning. *Educ. Psychol. Rev.*, 24(2):205–249.
- [Dweck 1999] Dweck, C. S. (1999). *Self-theories: Their role in motivation, personality, and development*. Self-theories: Their role in motivation, personality, and development. Psychology Press, New York, NY, US. Pages: xiii, 195.
- [Ferguson 2011] Ferguson, P. (2011). Student perceptions of quality feedback in teacher education. *Assess Eval High Educ*, 36(1):51–62.
- [Ferreira et al. 2020] Ferreira, M., Rolim, V., Mello, R. F., Lins, R. D., Chen, G., and Gašević, D. (2020). Towards automatic content analysis of social presence in transcripts of online discussions. In *Proceedings of the Tenth International Conference on Learning Analytics & Knowledge*, pages 141–150.
- [Ferreira-Mello et al. 2019] Ferreira-Mello, R., André, M., Pinheiro, A., Costa, E., and Romero, C. (2019). Text mining in education. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 9(6):e1332.
- [Friedman et al. 2001] Friedman, J., Hastie, T., and Tibshirani, R. (2001). *The elements of statistical learning*, volume 1. Springer series in statistics New York, NY, USA.
- [Hattie and Gan 2011] Hattie, J. and Gan, M. (2011). Instruction based on feedback. In *Handbook of research on learning and instruction*, pages 263–285. Routledge.
- [Hattie and Timperley 2007] Hattie, J. and Timperley, H. (2007). The power of feedback. *Review of educational research*, 77(1):81–112.
- [Laurillard 1993] Laurillard, D. (1993). *Rethinking University Teaching: Rethinking University Teaching: a Framework for the Effective Use of Educational Technology*. Routledge.
- [Liu et al. 2017] Liu, M., Li, Y., Xu, W., and Liu, L. (2017). Automated Essay Feedback Generation and Its Impact on Revision. *IEEE Trans. Learn. Technol.*, 10(4):502–513.
- [Ma et al. 2017] Ma, X., Wijewickrema, S., Zhou, S., Zhou, Y., Mhammedi, Z., O’Leary, S., and Bailey, J. (2017). Adversarial Generation of Real-time Feedback with Neural Networks for Simulation-based Training. In *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence*, pages 3763–3769, Melbourne, Australia. International Joint Conferences on Artificial Intelligence Organization.
- [Mani and Zhang 2003] Mani, I. and Zhang, I. (2003). knn approach to unbalanced data distributions: a case study involving information extraction. In *Proceedings of Workshop on Learning from Imbalanced Datasets*.
- [McNamara et al. 2014] McNamara, D. S., Graesser, A. C., McCarthy, P. M., and Cai, Z. (2014). *Automated evaluation of text and discourse with Coh-Metrix*. Cambridge University Press.
- [Neuendorf 2016] Neuendorf, K. A. (2016). *The content analysis guidebook*. Sage.

- [Nicol and Macfarlane-Dick 2006] Nicol, D. J. and Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. *Studies in higher education*, 31(2):199–218.
- [Osakwe et al. 2021] Osakwe, I., Chen, G., Whitelock-Wainwright, A., Gašević, D., Cavalcanti, A. P., and Mello, R. F. (2021). Towards automated content analysis of feedback: A multi-language study. In *Proceedings of the 14th International Conference on Educational Data Mining*.
- [Pan 2018] Pan, B. (2018). Application of XGBoost algorithm in hourly PM2.5 concentration prediction. *IOP Conference Series: Earth and Environmental Science*, 113:012127. Publisher: IOP Publishing.
- [Parikh et al. 2001] Parikh, A., McReelis, K., and Hodges, B. (2001). Student feedback in problem based learning: a survey of 103 final year students across five ontario medical schools. *Medical education*, 35(7):632–636.
- [Price et al. 2010] Price, M., Handley, K., Millar, J., and O'Donovan, B. (2010). Feedback : all that effort, but what is the effect? *Assess Eval High Educ*, 35(3):277–289.
- [Tausczik and Pennebaker 2010] Tausczik, Y. R. and Pennebaker, J. W. (2010). The psychological meaning of words: Liwc and computerized text analysis methods. *Journal of language and social psychology*, 29(1):24–54.
- [Villalón et al. 2008] Villalón, J., Kearney, P., Calvo, R. A., and Reimann, P. (2008). Glosser: Enhanced Feedback for Student Writing Tasks. In *2008 Eighth IEEE International Conference on Advanced Learning Technologies*, pages 454–458, Santander, Cantabria, Spain. IEEE.
- [Weaver 2006] Weaver, M. R. (2006). Do students value feedback? Student perceptions of tutors' written responses. *Assess Eval High Educ*, 31(3):379–394.
- [Xiao et al. 2017] Xiao, Z., Wang, Y., Fu, K., and Wu, F. (2017). Identifying Different Transportation Modes from Trajectory Data Using Tree-Based Ensemble Classifiers. *ISPRS International Journal of Geo-Information*, 6(2):57. Number: 2 Publisher: Multidisciplinary Digital Publishing Institute.