



José Rodrigues de Lima Neto

Tutoria: Plataforma para suporte à correção de atividades e envio de feedback personalizado

Recife

2023

José Rodrigues de Lima Neto

Tutoria: Plataforma para suporte à correção de atividades e envio de feedback personalizado

Monografia apresentada ao Curso de Bacharelado em Ciências da Computação da Universidade Federal Rural de Pernambuco, como requisito parcial para obtenção do título de Bacharel em Ciências da Computação.

Universidade Federal Rural de Pernambuco – UFRPE

Departamento de Computação

Curso de Bacharelado em Ciências da Computação

Orientador: Rafael Ferreira Leite de Mello

Recife

2023

Dados Internacionais de Catalogação na Publicação
Universidade Federal Rural de Pernambuco
Sistema Integrado de Bibliotecas
Gerada automaticamente, mediante os dados fornecidos pelo(a) autor(a)

R696t Neto, José Rodrigues de Lima
Tutoria: Plataforma para suporte à correção de atividades e envio de feedback personalizado / José Rodrigues de Lima Neto. - 2023.
27 f. : il.

Orientador: Rafael Ferreira Leite de Mello.
Inclui referências.

Trabalho de Conclusão de Curso (Graduação) - Universidade Federal Rural de Pernambuco,
Bacharelado em Ciência da Computação, Recife, 2023.

1. feedback. 2. plataforma Tutoria. 3. inteligência artificial. 4. usabilidade. 5. correção de questões discursivas. I. Mello, Rafael Ferreira Leite de, orient. II. Título

CDD 004



**MINISTÉRIO DA EDUCAÇÃO E DO ESPORTO
UNIVERSIDADE FEDERAL RURAL DE PERNAMBUCO (UFRPE)
BACHARELADO EM CIÊNCIA DA COMPUTAÇÃO**

<http://www.bcc.ufrpe.br>

FICHA DE APROVAÇÃO DO TRABALHO DE CONCLUSÃO DE CURSO

Trabalho defendido por José Rodrigues de Lima Neto às 17 horas e 0 minutos do dia 19 de abril de 2023, no link <https://meet.google.com/pjs-bnbw-jte>, como requisito para conclusão do curso de Bacharelado em Ciência da Computação da Universidade Federal Rural de Pernambuco, intitulado “Tutoria: Plataforma para suporte à correção de atividades e envio de feedback personalizado”, orientado por Rafael Ferreira Leite de Mello e aprovado pela seguinte banca examinadora:

Rafael Ferreira Leite de Mello
DC/UFRPE

André Câmara Alves do Nascimento
DC/UFRPE

À memória de meu avô, José Rodrigues de Lima, eterno vovô Uza.

Agradecimentos

Primeiramente, gostaria de expressar minha gratidão a Deus, o Arquiteto Supremo do Universo, pela minha existência, amor e pela força que me foi concedida para superar todas as barreiras físicas, psicológicas e emocionais durante o curso.

Agradeço especialmente à minha mãe, Lidiane Lemos, que mesmo diante de tantas dificuldades financeiras, lutou incansavelmente para que eu pudesse estudar, ter acesso a uma boa alimentação, roupas novas e todos os materiais necessários para o meu aprendizado.

Não posso deixar de mencionar meu pai, José Soares, que adquiriu o primeiro computador da casa para que minha mãe pudesse fazer um curso de informática, dando origem ao meu fascínio pela computação. Ele também foi meu motorista pessoal durante o curso e contribuiu financeiramente para meus gastos na faculdade.

Agradeço à minha irmã, Gislainy Thais, que foi a primeira pessoa da minha família a se graduar e sempre me incentivou a ingressar na universidade, além de oferecer suporte emocional e físico inestimável.

Sou muito grato ao meu orientador Rafael Mello, que além de ser um excelente orientador, foi um amigo que sempre esteve presente para me aconselhar e encorajar. Ele é um modelo a ser seguido, tanto profissionalmente quanto pessoalmente, sendo o principal responsável pela escolha da inteligência artificial como meu foco de aprendizagem no curso e tema do meu TCC.

Não posso deixar de mencionar meu professor Péricles Miranda, a quem tenho uma grande gratidão por me oferecer a oportunidade de participar da minha primeira Iniciação Científica e pela minha primeira experiência no mercado de trabalho, que está diretamente relacionado ao emprego que tenho hoje.

Gostaria de também agradecer à minha namorada, Giulia Maia, que conheci durante o curso. Ela tem sido uma grande companheira, me dando suporte emocional e encorajamento ao longo da jornada acadêmica. Sua presença em minha vida tem sido fundamental para meu crescimento e desenvolvimento pessoal e profissional.

Também gostaria de agradecer a todos os amigos que fiz durante o curso que se tornaram irmãos e irmãs ao longo do bacharelado. Estiveram sempre presentes em cada uma das etapas de evolução que tive e pudemos trocar experiências de aprendizado em disciplinas e no mundo profissional.

Sou grato a todos que me apoiaram durante essa jornada, que me ajudaram a superar desafios e a crescer como pessoa e profissional.

*“Qualquer um pode ser vencedor, desde que não haja nenhum concorrente.”
(Falcão, Marcondes)*

Resumo

O feedback dado por professores a estudantes sobre atividades avaliativas é fundamental para a construção do conhecimento e compreensão acerca de sua trajetória de aprendizagem. Entretanto, frequentemente professores não conseguem fornecer feedback de qualidade e em tempo hábil, devido à sobrecarga de trabalho e à falta de tempo, gerando frustração para todos os envolvidos. Neste contexto, a plataforma Tutoria tem o objetivo de apoiar professores na correção de atividades avaliativas de maneira mais rápida, mantendo a qualidade e a personalização. Para isso, Tutoria utiliza técnicas de inteligência artificial para dar suporte na correção de questões discursivas, priorizando a usabilidade.

Palavras-chave: feedback, plataforma Tutoria, inteligência artificial, usabilidade, correção de questões discursivas.

Abstract

The feedback message provided by instructors to students about assessment activities is fundamental for building knowledge and understanding about their learning trajectory. However, instructors are often unable to provide quality feedback promptly due to work overload and lack of time, generating frustration for everyone involved. In this context, the Tutoria platform aims to support instructors in the faster assessment of academic activities while maintaining quality and personalization. Tutoria prioritizes usability, discursive questions correction support, and the use of artificial intelligence techniques.

Keywords: feedback, Tutoria platform, artificial intelligence, usability, correction of discursive questions.

Lista de ilustrações

Figura 1 – Tela inicial das atividades dos alunos.	14
Figura 2 – Tela de correção de questão aberta.	15
Figura 3 – Fluxo de sistema de recomendação de tag proposto.	16
Figura 4 – Tela de correção de questão múltipla escolha.	17
Figura 5 – Tela de elaboração de devolutiva.	18
Figura 6 – Dashboard com estatísticas da turma.	18

Lista de tabelas

Tabela 1 – Comparação entre as ferramentas relacionadas ao Tutoria	13
Tabela 2 – Distribuição de tags por questão.	19
Tabela 3 – Resultados de cada algoritmo na recomendação de tag.	23

Sumário

	Lista de ilustrações	7
1	INTRODUÇÃO	10
2	TRABALHOS RELACIONADOS	12
3	PLATAFORMA TUTORIA	14
4	METODOLOGIA	19
4.1	Arquitetura de curso e dados	19
4.2	Processamento de texto e extração de características	20
4.2.1	Características do TF-IDF	20
4.2.2	BERT	20
4.3	Medidas de Semelhança e Avaliação	20
5	RESULTADOS	23
6	APLICAÇÃO INICIAL	24
7	CONSIDERAÇÕES FINAIS	25
	REFERÊNCIAS	26

1 Introdução

Em contextos formais de ensino, a aprendizagem dos estudantes é verificada por meio de atividades avaliativas, geralmente entregues por escrito. Nesses contextos, os professores são a principal fonte de feedback, orientação e validação para os estudantes em sua trajetória educacional. O feedback dado pelo professor por meio da correção das atividades é fundamental para a construção do conhecimento pelos estudantes e para a autorregulação de sua aprendizagem (WIGGINS, 1998).

Especialmente no ensino superior, problemas como a falta de tempo, o excesso de estudantes por turmas e a sobrecarga de trabalho (BOUD; MOLLOY, 2013), dificultam o fornecimento de feedback por parte dos professores. Estes problemas se intensificaram ainda mais com a massificação da modalidade de ensino remoto, adotada emergencialmente como alternativa educacional face às condições sanitárias restritivas impostas pela pandemia da COVID-19 nos anos de 2020 e 2021. Tal sobrecarga frequentemente leva a atrasos na devolutiva das atividades dos estudantes. Há casos em que apenas a nota ou conceito é considerado devolutiva, o que não favorece a compreensão do estudante sobre as lacunas e as consequentes ações que poderia tomar para melhorar a sua aprendizagem (CAVALCANTI et al., 2021).

Além disso, dar feedback por escrito apresenta outros desafios, como a linguagem utilizada pelo professor, que muitas vezes é demasiadamente formal, e o conteúdo das devolutivas, que frequentemente não se transpõe claramente em ações que o estudante deva tomar para melhorar sua aprendizagem e desempenho (HIGGINS; HARTLEY; SKELTON, 2001; IVANIC; CLARK; RIMMERSHAW, 2000). Um feedback de qualidade deve ser personalizado; apresentar sugestões objetivas e específicas; mostrar as lacunas de aprendizagem em relação ao que é esperado do estudante naquela atividade; evitar um tom excessivamente crítico; estimular o diálogo; contemplar pontos positivos; e motivar os estudantes a persistirem e melhorarem (FREEMAN; LEWIS, 2016; NICOL; MACFARLANE-DICK, 2006). No entanto, nem sempre o professor consegue fornecer um feedback com estas características, gerando insatisfação e frustração mútuo (BOUD; MOLLOY, 2013): professores por terem consciência de que não estão fornecendo um feedback de qualidade; e estudantes, por não terem um retorno claro sobre seus erros que possa orientar seus estudos visando um melhor desempenho.

Além disto, a disseminação dos ambientes virtuais de aprendizagem (AVA) como Moodle e Google Classroom, não somente em cursos a distância mas também nos presenciais, têm aumentado consideravelmente a quantidade de atividades avaliativas en-

tregues no formato digital. Isso amplia as possibilidades de ferramentas de software que possam ajudar os professores no processo de correção de atividades. Porém, a maioria das ferramentas desenvolvidas foca a correção automática de questões objetivas e o envio de mensagens automáticas sem a interferência do professor. Este comportamento limita o escopo e possibilidades da ferramenta e impede a realização de devolutivas personalizadas (CAVALCANTI et al., 2021).

Diante deste contexto, este artigo apresenta a plataforma Tutoria, cujo principal diferencial é auxiliar professores a corrigir questões discursivas e elaborar uma devolutiva escrita, personalizada e de qualidade. A ferramenta tem sido desenvolvida com foco no ensino superior, embora não se restrinja a este nível, já que sua estrutura genérica permite a adoção também em escolas ou outros contextos de ensino formal. Tutoria pode ser aplicada em quaisquer disciplinas e áreas do conhecimento, embora o seu foco seja em correção de textos, não havendo, por exemplo, um suporte específico para correção de algoritmos escritos em linguagens de programação, ou cálculos matemáticos. Neste artigo, apresentaremos resultados iniciais da utilização do Tutoria, na prática.

2 Trabalhos relacionados

Para entender como os produtos existentes atendem às necessidades dos professores na tarefa de corrigir atividades e dar feedback, analisamos as plataformas mais relacionadas que encontramos disponíveis no mercado.

AvaliA (AVALIA, 2019) é uma solução de software que permite a criação automática de atividades, a partir de um banco de questões com mais de 90.000 opções e correção automática de folhas de respostas. Ele suporta a escolha de questões com níveis diferentes de dificuldade. Além disso, gera diferentes tipos de relatórios para administradores, coordenadores, professores e alunos, e oferece um ambiente controlado para aplicação supervisionada de exames. Ao usar o AvaliA, os professores devem escolher perguntas predefinidas do banco, todas de múltipla escolha. Não há suporte para gerar feedback que vá além do “certo ou errado” e da nota.

O OnTask (PARDO et al., 2018) é uma ferramenta de software que reúne e avalia dados sobre as atividades dos alunos e permite que os professores criem feedback personalizado com sugestões sobre estratégias de aprendizado. Essa ferramenta ajuda os professores a criar feedback personalizado, adaptando a mensagem para cada aluno usando um esquema de regras *if/else*, com atributos. Ele aprimora a capacidade do instrutor de fornecer feedback para grandes grupos de alunos criando apenas uma mensagem (PARDO et al., 2018). No entanto, a interface é bastante técnica, exigindo conhecimentos básicos na manipulação de variáveis. Além disso, restringe-se à criação de feedback com base em uma análise geral do progresso da aula e não oferece suporte para correção de tarefas. Embora o OnTask tenha se mostrado eficaz para melhorar o desempenho e a motivação dos alunos, ele ainda não foi amplamente adotado, principalmente devido à necessidade de habilidades específicas (por exemplo, conhecimento básico de linguagem de programação) dos professores para produzir a mensagem de feedback (TSAI et al., 2021).

Com Gradescope (SINGH et al., 2017), os professores podem corrigir diferentes tipos de tarefas (projetos, planilhas, questionários, etc.), incluindo respostas dadas em texto ou múltipla escolha. Os professores podem avaliar e escrever comentários para cada item da tarefa, usando rubricas que podem ser atualizadas de forma flexível. A ferramenta gera relatórios sobre o desempenho dos alunos. O Gradescope também oferece classificação assistida por Inteligência Artificial, que reconhece respostas semelhantes e permite classificar os grupos, tornando o processo mais eficiente.

A tabela 1 compara as ferramentas de acordo com características importantes para auxiliar os professores no fornecimento de feedback de qualidade com eficiência.

Tabela 1 – Comparação entre as ferramentas relacionadas ao Tutoria

	OnTask	AvaliA	Gradescope	Tutoria
Correção de atividades	x	✓	✓	✓
Elaboração de feedback	✓	x		✓
Diminuição do tempo de correção	x	✓	✓	✓
Autonomia dos professores	✓	x	✓	✓
Integração com ambientes virtuais	x	x	✓	✓
Suporte para elaboração de mensagens de feedback de qualidade	x	x	x	✓
Reaproveitamento inteligente de feedbacks	x	x	x	✓

O Gradescope abrange todos eles, exceto fornecer suporte para escrever mensagens de feedback de qualidade, com alto impacto na motivação dos alunos e na aprendizagem autorregulada (NICOL; MACFARLANE-DICK, 2006). Além disso, é importante observar que AvaliA e Gradescope são softwares proprietários, enquanto OnTask é uma ferramenta de código aberto (mas requer conhecimento técnico específico para ser instanciado em uma instituição).

3 Plataforma Tutoria

A plataforma Tutoria¹ oferece suporte para correção de tarefas escritas e elaboração de feedback direcionado para o usuário. Para isso, os professores importam as atividades do ambiente virtual de aprendizagem (por exemplo, Google Sala de Aula ou Moodle). Após importar as respostas, o professor pode optar por realizar a correção por pergunta ou aluno (Figura 1).

Atualmente a plataforma Tutoria oferece suporte à correção de questões abertas e múltipla escolha. Para corrigir uma questão aberta (Figura 2) o professor pode criar um conjunto de tags, relacionadas a erros e acertos, que contém um feedback próprio. Essas tags podem ser reaproveitadas para corrigir atividades de diferentes alunos, assim, o professor consegue facilmente reutilizar a mensagem de feedback escrita. Além disto, o Tutoria utiliza algoritmos de inteligência artificial para recomendar tags que já foram criadas para correções futuras, quando encontra textos similares, otimizando ainda mais o trabalho do professor.

A inteligência artificial atua no fluxo de correção de atividades abertas como mostrado na Figura 3, no primeiro passo o professor recebe a resposta do Estudante 1, seleciona os trechos de erro ou acerto do aluno e gera um feedback para cada tag, como representado no passo 2. Procedendo à correção para o Estudante 2, apresen-

¹ <https://tutor-ia.com/>

The screenshot displays the Tutoria platform's interface for activity correction. At the top, there is a navigation bar with the logo 'TUTORIA' and links for 'Configurações', 'Ajuda', 'Suporte', and 'Minha conta'. Below this, the main content area is titled 'Visão geral' and shows 'Visualizando informações sobre a atividade Brasil Colonial'. A search bar contains the text 'Ex.: história do Brasil'. There are two tabs: 'Atividades para corrigir' (selected) and 'Atividades já corrigidas'. Below the tabs, there are buttons for 'POR QUESTÃO' and 'POR ESTUDANTE'. A row of buttons labeled 'A', 'B', 'C', 'D', and 'E' is visible. The main content is divided into two sections for 'Estudante AA' and 'Estudante AB'. Each student's section contains three question cards. The first card in each section has a 'RETOMAR CORREÇÃO' button. The questions are: 'Questão 1: Descreva sobre o Brasil colônia, evidenciando o relacionamento entre índios e portugueses [...]', 'Questão 2: Em novembro de 1807, a Família Real embarcou rumo ao Brasil, pois Portugal [...]', and 'Questão 3: Identifique o papel do Partido Brasileiro no processo de independência do País, bem [...]'.

Figura 1 – Tela inicial das atividades dos alunos.

The screenshot displays the 'Minha correção' (My correction) interface. At the top, there's a navigation bar with 'TUTORIA', 'Configurações', 'Ajuda', 'Suporte', and 'Minha conta'. The main content area shows the question: '1. Qual a importância do dia 7 de setembro para o Brasil?'. The user's response is 'ACERTO PARCIAL'. A feedback message is shown: '07 de setembro é uma das datas comemorativas mais importantes do Brasil, justamente por abrigar um dos principais acontecimentos da nossa história: a proclamação da república'. Below this, there's a section for 'Marcadores da questão' (Question markers) with a table showing 'ERRO', 'ACERTO', and 'ACERTO PARCIAL'. A text input field for 'Nome do marcador' (Marker name) is present, along with a note to 'Insira aqui uma explicação para o marcador selecionado' (Insert an explanation for the selected marker). A 'Comentário geral' (General comment) section is also visible. On the right, a 'Progresso de correção' (Correction progress) sidebar shows '81,6% Você já corrigiu 49/60 das provas' and '100% Você já corrigiu 4/4 respostas da Questão 1'. A list of markers used in the question is shown: '07 de Setembro', 'a proclamação da república', and 'lorem ipsum'. Navigation buttons 'VOLTAR' and 'PRÓXIMO' are at the bottom.

Figura 2 – Tela de correção de questão aberta.

tado no passo 3, o algoritmo inteligente recomenda ao professor um possível erro ou acerto no trecho da resposta do aluno baseado nas tags marcadas na resposta do primeiro estudante, essa recomendação pode ou não ser aceita pelo usuário, ao passo 4 o professor aceita a recomendação e, automaticamente, o conteúdo do feedback já redigido pelo professor é reaproveitado para a resposta do estudante atual.

O Tutoria também dá suporte a correção e criação de mensagem de feedback em questões de múltipla escolha. A Figura 4 apresenta os detalhes desta opção. Nela o professor escolhe a questão correta, com a opção de escrever uma mensagem explicando porque cada alternativa está correta ou errada. Nesta opção, não há recomendação de tags já que a questão é objetiva e todos os alunos recebem a mesma pontuação dependendo da pontuação atribuída a cada alternativa escolhida. Apesar de não ter a utilização de tags, é possível incluir um comentário geral, assim como ocorre nas questões abertas, mas que neste caso o comentário é exibido para todos os estudantes. Todavia, mesmo este comentário geral pode ser customizado para cada estudante, posteriormente na etapa de envio da devolutiva.

Após a realização da correção, os professores podem criar um modelo de feedback para ser usado para enviar a devolutiva para o aluno. A Figura 5 apresenta os aspectos relevantes deste modelo: (i) para personalizar a mensagem o professor

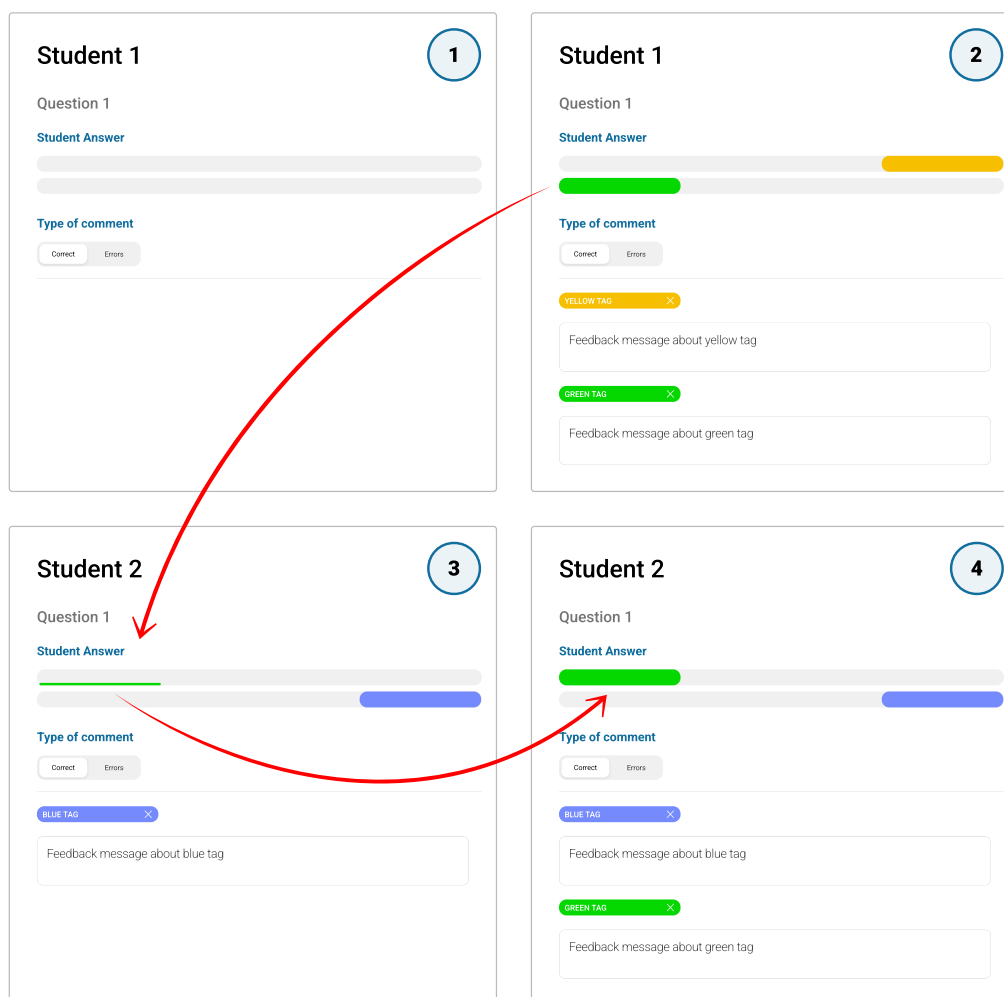


Figura 3 – Fluxo de sistema de recomendação de tag proposto.

pode utilizar um conjunto de atributos, como nome ou nota do aluno, que o sistema vai recuperar automaticamente para incluir na mensagem final; (ii) cada questão corrigida gera um bloco de feedback específico relacionado às respostas dos alunos. Em outras palavras, todas as mensagens cadastradas nas tags que o professor usou em cada questão são selecionadas para os alunos e elas compõem a devolutiva final. Vale destacar que o professor pode criar quantos blocos textuais achar relevante para conectar os diferentes blocos de feedback por questão.

Por fim, a Figura 6 apresenta um dashboard que está sendo desenvolvido e em breve estará integrado ao Tutoria. Ele apresenta informações sobre as tags de erros e acertos que são mais frequentes em cada questão e um acompanhamento geral do desempenho dos estudantes em um curso. Essas informações podem ser usadas pelo professor para tomar decisão sobre possíveis mudanças metodológicas na disciplina.

TUTORIA
Configurações Ajuda Suporte
Minha conta ▾

👤
🗄️
✍️
💬
🔗
📍
🔄
🕒

Minha correção

Você está corrigindo a atividade **Brasil Colonial**

1. Qual a importância do dia 7 de setembro para o Brasil?

a) Lorem ipsum dolor sit amet, consectetur adipiscing elit. 0,5 CORRETO ▾

b) Praesent congue leo non odio efficitur, et viverra nisi dignissim. 0,0 INCORRETO ▾

c) Aliquam et ipsum auctor, commodo purus a, efficitur felis. 0,0 INCORRETO ▾

d) Curabitur vel metus scelerisque metus cursus feugiat. 0,0 INCORRETO ▾

e) Duis ornare diam in bibendum iaculis. 0,25 PARCIALMENTE CORRETO ▾

Jusificar alternativas

A B **C** D E

Quis eu viverra maecenas ac. Ipsum cum donec hendrerit gravida tempus eget auctor. Diam erat id convallis eros arcu, semper dui purus. Non sit interdum imperdiet accumsan senectus neque molestie odio amet. Et lectus vitae est sodales pulvinar elementum diam, risus. Voluptat tellus malesuada velit lobortis porttitor eu. Pellentesque non nulla fringilla non enim suscipit.

c) Aliquam et ipsum auctor, commodo purus a, efficitur felis.

Quis eu viverra maecenas ac. Ipsum cum donec hendrerit gravida tempus eget auctor. Diam erat id convallis eros arcu, semper dui purus. Non sit interdum imperdiet accumsan senectus neque molestie odio amet. Et lectus vitae est sodales pulvinar elementum diam, risus. Voluptat tellus malesuada velit lobortis porttitor eu. Pellentesque non nulla fringilla non enim suscipit.

Comentário geral ⓘ

Nota para a questão

0,0 ponto

VOLTAR
PRÓXIMO

Aprenda a usar o Tutoria

Progresso de correção

81,6% Você já corrigiu 49/60 das provas

100% Você já corrigiu 4/4 respostas da Questão 1

26,6% Você já corrigiu 4/15 respostas do estudante

MARCOS LINS

Os principais marcadores dessa questão são:

- 07 de Setembro
- a proclamação da república
- lorem ipsum

Figura 4 – Tela de correção de questão múltipla escolha.

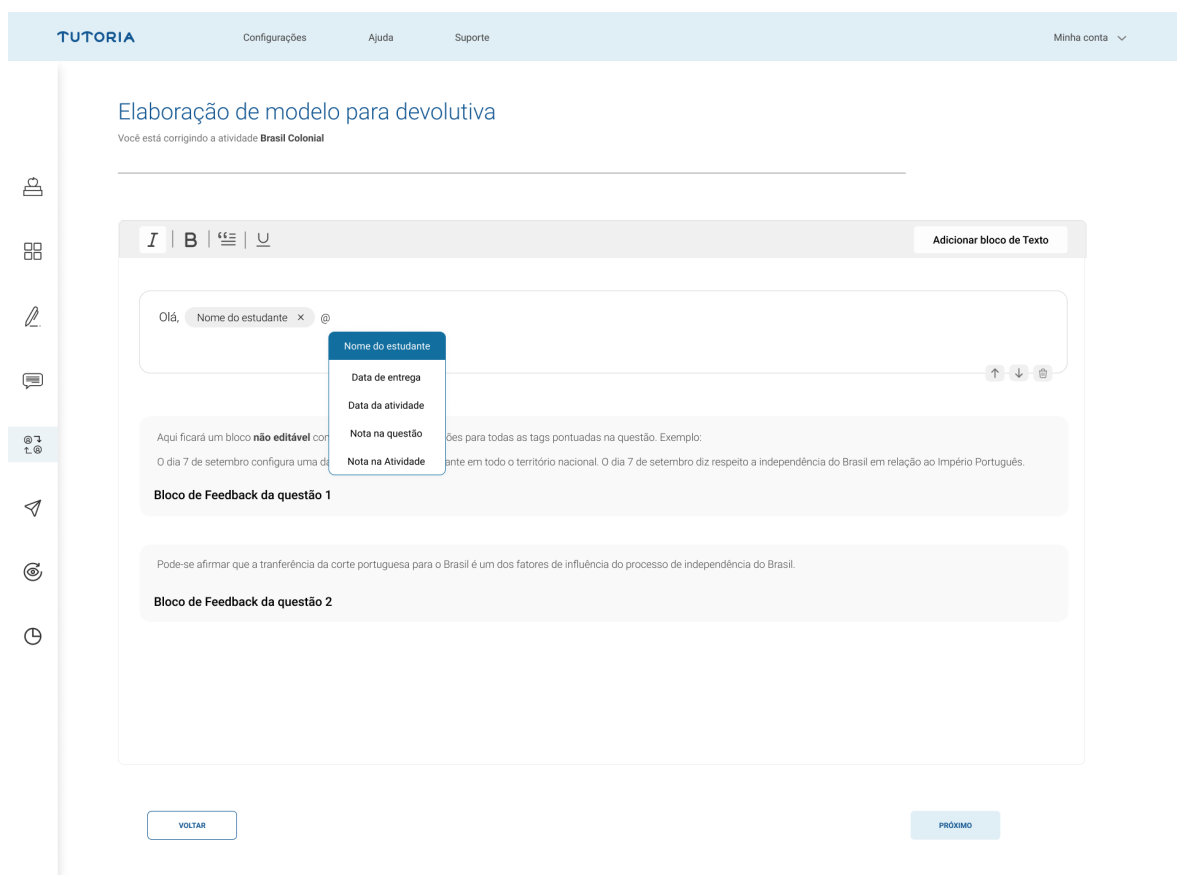


Figura 5 – Tela de elaboração de devolutiva.

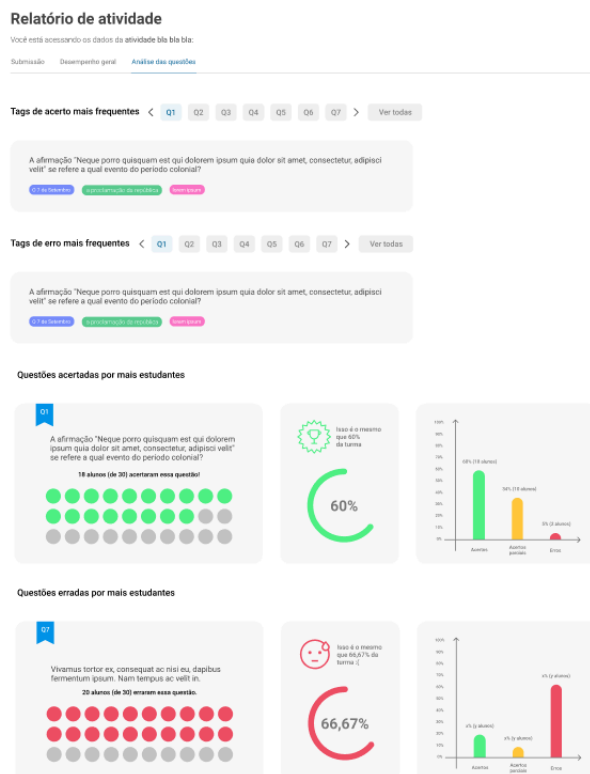


Figura 6 – Dashboard com estatísticas da turma.

4 Metodologia

4.1 Arquitetura de curso e dados

Os dados educacionais utilizados para a validação da inteligência artificial correspondem a um exercício extraído de um curso de graduação totalmente online sobre Informática Básica, que explora temas relacionados a hardware, software, redes, sistema operacional, entre outros.

Este curso incluiu uma série de vídeos instrutivos sobre diferentes tópicos usados em combinação com tarefas online contendo perguntas de múltipla escolha e abertas. A cada duas semanas, novos vídeos e tarefas eram fornecidos aos alunos. Estes trabalhos representaram 50% da nota final. Na oferta do curso analisado, um total de 47 alunos responderam à primeira tarefa, contendo cinco questões abertas.

O instrutor deste curso, com formação em ciência da computação, concordou em usar a plataforma Tutoria para avaliar as respostas abertas sem o sistema de recomendação de tags, a fim de gerar as tags para cada resposta manualmente e produzir o padrão ouro neste estudo. Assim, neste estudo avaliamos o sistema de recomendação com base nas tags incluídas por um instrutor. A tabela 2 apresenta os detalhes do número de tags dividido por cada questão. Mostra o (1) número de respostas dos alunos; (2) número de tags exclusivas que o instrutor incluiu; (3) número total de tags, incluindo a repetição da mesma tag para respostas diferentes; e (4) número máximo de tags exclusivas que podem ser sugeridas se o sistema recomendar todas as tags exclusivas para todas as respostas dos alunos. Neste experimento avaliamos respostas com 100-200 palavras.

Tabela 2 – Distribuição de tags por questão.

Questão	Respostas	Tags únicas	Número total de tags	Recomendação potencial
Q1	33	4	35	132
Q2	33	4	42	132
Q3	47	10	98	470
Q4	47	3	63	141
Q5	47	4	37	188
Total	207	25	275	1063

4.2 Processamento de texto e extração de características

Como primeiro passo de nossa análise, usamos técnicas de Processamento de Linguagem Natural (PLN) para processar o texto e extrair recursos. As medidas de similaridade adotadas neste estudo precisam ser aplicadas apenas às palavras. Portanto, removemos pontuação, vários espaços e caracteres Unicode. Além disso, também aplicamos métodos que se baseiam em pontuações de Term-Frequency Inverse Document Frequency (TF-IDF) e Bidirectional Encoder Representations from Transformers (BERT).

4.2.1 Características do TF-IDF

TF-IDF é uma das abordagens mais utilizadas em modelos de mineração de texto para extrair características de textos (MANNING; SCHUTZE, 1999). Esse algoritmo converte documentos textuais (por exemplo, respostas dos alunos) em um vetor que consiste na contagem de termos (MANNING; SCHUTZE, 1999), neste caso os valores TF-IDF. O presente estudo adotou a técnica tradicional de TF-IDF (MANNING; SCHUTZE, 1999).

4.2.2 BERT

Também adotamos o BERT para incluir uma abordagem de aprendizado profundo estado da arte na comparação. BERT é uma abordagem de incorporação de palavras que considera o contexto de cada palavra, o que demonstrou aumentar o desempenho em vários aplicativos de PLN (DEVLIN et al., 2019). Estudos anteriores mostraram o potencial do uso do BERT em sistemas automatizados de classificação de respostas curtas (BONTHU; SREE; PRASAD, 2021). É importante mencionar que não fizemos nenhuma preparação de dados em nosso conjunto de dados antes de usar o BERT, como sugerido pelos estudos anteriores (DEVLIN et al., 2019; BONTHU; SREE; PRASAD, 2021).

4.3 Medidas de Semelhança e Avaliação

As medidas de similaridade avaliadas neste artigo são compostas por métodos estatísticos para realizar correspondência de strings, correspondência de palavras e a abordagem de aprendizado profundo usando BERT (DEVLIN et al., 2019). O resultado de cada medida de similaridade é um número entre 0 e 1, onde 1 significa a maior similaridade. Com base nos estudos anteriores, decidimos usar um limite de similaridade igual a 0,7 para definir o texto que deveria receber uma recomendação de

tag (MARIN, 2004; CUTRONE; CHANG, 2010; SIDDIQI; HARRISON; SIDDIQI, 2010; BONTHU; SREE; PRASAD, 2021).

O primeiro grupo avaliado é baseado em medidas de similaridade de correspondência de strings que buscam encontrar substrings com sobreposições a nível de caractere. O método mais conhecido é a distância Levenshtein, que conta o número de modificações que devem ser feitas para mudar uma string para a outra (YUJIAN; BO, 2007). A lista a seguir descreve as medidas que foram avaliadas neste grupo.

Levenshtein: Usamos a implementação clássica da distância de Levenshtein.

Partial ratio: Este algoritmo realiza a correspondência de similaridade da string mais curta com todas as substrings de mesmo comprimento.

Token Sort Ratio: Essa métrica realiza um processo de tokenização para limpar a string antes do cálculo da distância de Levenshtein.

Partial Token Sort Ratio: Essa medida executa um processo de tokenização para limpar a string antes da correspondência final usando a distância Levenshtein.

Token Set Ratio: Esse método usa o Token Sort Ratio, mas também adiciona um processo de remoção de stopword antes da correspondência final com a distância Levenshtein.

Partial Token Set Ratio: É semelhante ao Token Set Ratio, mas usa a semelhança de proporção parcial em vez da distância de Levenshtein.

Fuzzy Search: Utiliza a distância Levenshtein para procurar um grupo de substrings semelhantes em vez de avaliar o texto inteiro.

Edit Distance: Esta medida usa a distância Jaro-Winkler (WANG; QIN; WANG, 2017) para calcular a semelhança final.

Rapidfuzz: Esta é uma versão mais rápida da similaridade Edit Distance.

O segundo grupo é baseado na correspondência de palavras aplicando uma composição diferente de n-gram. Ele usa as pontuações do TF-IDF para vetorizar os textos analisados, computando as semelhanças usando:

1-gram: Compara a similaridade de cada palavra nos dois textos.

2-gram: Compara a similaridade de cada par de palavras nos dois textos.

3-gram: Compara a similaridade de cada segmento de três palavras em ambos os textos.

4-gram: Compara a similaridade de cada segmento de quatro palavras nos dois textos.

n-gram: Ele usa todas as semelhanças anteriores (1, 2, 3 e 4-gram) para calcular a pontuação final.

Por fim, também avaliamos o desempenho do modelo BERT neste estudo. Ao contrário dos métodos de correspondência de palavras, o BERT encapsula um vetor por palavra, não por frase. Significa que a similaridade, neste caso, compara duas matrizes. A ideia principal é ter informações semânticas sobre as palavras que estão sendo comparadas. Porém, aumenta o tempo para se ter o resultado final.

Para avaliar o desempenho das medidas de similaridade propostas, adotamos as medidas tradicionais de aprendizado de máquina Precision (P), Recall (R) e F1-score, amplamente utilizadas neste contexto (MANNING; SCHUTZE, 1999). Em suma, a precisão mediu o número de etiquetas recomendadas que o instrutor aceitou como uma etiqueta correta; e o recall avaliou o número de tags relevantes para uma resposta específica que o sistema não recomendou. O F1-score é a média harmônica entre precisão e recall. Além disso, também avaliamos o desempenho em termos de tempo para recomendar a etiqueta, pois é uma questão crítica para o uso prático da abordagem proposta.

5 Resultados

Os resultados apresentados nesta seção mostram o desempenho de cada medida de similaridade para a recomendação de 25 tags para respostas de 207 alunos (ver seção 4.1 para mais detalhes). Na pior das hipóteses, o sistema recomendaria 1.063 tags, o número total de tags para cada pergunta multiplicado pelo número de respostas. A tabela 3 apresenta os resultados de cada algoritmo de similaridade que foi avaliado usando precisão, recall, F1-Score e o tempo para executar a recomendação para todas as respostas em segundos.

Em geral, todos os algoritmos alcançaram bons resultados em termos de precisão. Na pior das hipóteses, a similaridade Levenshtein atingiu 0,80. Isso significa que os algoritmos conseguiram recomendar tags para os instrutores corretamente. Em contraste, algoritmos múltiplos obtiveram recall inferior a 0,50. Em outras palavras, eles sugeriram um pequeno número de tags, o que significa que o sistema não recomendou tags para a maioria das questões e o instrutor teve que fazê-lo manualmente, mantendo uma carga de trabalho semelhante à não utilização da abordagem proposta.

Três medidas de similaridade alcançaram F1-Score superior a 0,80: Partial Token Set Ratio, TF-IDF 1-gram e BERT. Esses algoritmos alcançaram um equilíbrio entre precisão e revocação, e foram os algoritmos mais adequados para esta tarefa. No entanto, o BERT foi o mais lento, levando 93,28 segundos para realizar as previsões de tags, enquanto o Parcial Token Set Ratio fez as mesmas recomendações em 0,08 segundos.

Tabela 3 – Resultados de cada algoritmo na recomendação de tag.

#	Algoritmo de Similaridade	Precision	Recall	F1-Score	Tempo Médio	Tempo Mediano
1	Levenshtein	0.80	0.01	0.01	00.02	00.03
2	Partial Ratio	0.98	0.32	0.48	00.06	00.29
3	Token Sort Ratio	0.94	0.02	0.03	00.05	00.07
4	Token Set Ratio	0.97	0.43	0.59	00.05	00.06
5	Partial Token Set Ratio	0.91	0.88	0.89	00.08	00.20
6	Partial Token Sort Ratio	0.96	0.25	0.39	00.07	00.25
7	Fuzzy Search	0.90	0.42	0.57	01.22	26.95
8	Edit Distance	0.93	0.62	0.74	01.03	01.04
9	Rapidfuzz	0.93	0.60	0.72	00.86	00.87
10	TFIDF 1-gram	0.90	0.74	0.81	05.10	05.10
11	TFIDF 2-gram	0.98	0.06	0.10	04.84	04.83
12	TFIDF 3-gram	0.98	0.02	0.03	04.78	04.75
13	TFIDF 4-gram	0.96	0.01	0.02	04.71	04.62
14	TFIDF n-gram	0.94	0.19	0.31	05.27	05.39
15	BERT	0.89	0.79	0.83	93.28	107.86

6 Aplicação Inicial

A plataforma Tutoria está disponível para ser acessada a qualquer momento ¹. Foi realizada uma avaliação inicial da plataforma com professores do departamento de computação de uma universidade pública brasileira. Durante uma chamada do Google Meet, que durou cerca de uma hora, 10 professores foram convidados a realizar a correção em uma atividade de informática básica que foi criada para esta avaliação. Os professores tiveram que realizar todo o procedimento requerido pela ferramenta, desde a importação das questões até o envio do feedback. Contudo, não foram dadas instruções sobre como a correção é realizada ou como o feedback pode ser construído e enviado com [omitido], para que fossem avaliados o uso autônomo e a curva de aprendizado dos novos usuários do Tutoria.

Ao final das tarefas, os instrutores foram incentivados a comentar oralmente sobre problemas encontrados ou sugestões que tivessem para a plataforma. A ligação foi gravada e todas as contribuições foram posteriormente discutidas pela equipe para decidir como fazer ajustes no produto. Os instrutores também foram solicitados a preencher um formulário com uma escala de Likert com 5 pontos, baseado na System Usability Scale (SUS) (GRIER et al., 2013). Como resultado, a plataforma obteve 73 pontos, que é considerado um bom indicativo de usabilidade.

¹ <https://tutor-ia.com/>

7 Considerações Finais

A plataforma Tutoria visa a auxiliar professores na correção e elaboração de mensagens de feedback de qualidade. Para isso, utiliza técnicas de inteligência artificial e conceitos de interação homem-computador para criar um sistema baseado em tags, capaz de otimizar o trabalho do professor. Esta ferramenta já recebeu vários prêmios e reconhecimentos recentes [omitidos], mostrando o potencial de sua aplicação.

Atualmente a plataforma se encontra num estágio avançado de desenvolvimento, mas seus principais serviços já estão disponíveis para serem usados através do site. Embora a aplicação da plataforma ainda possua um número restrito de usuários, ela está sendo desenvolvida para ser escalável. Ademais, professores de diferentes instituições estão participando de um piloto com o Tutoria com foco em testes na plataforma.

Por fim, vale destacar que o principal objetivo do Tutoria é o educacional, provendo funcionalidades que auxiliam o professor, com impacto direto na aprendizagem do aluno. Sendo assim, todas as técnicas e abordagens utilizadas no seu desenvolvimento objetivam contribuir para o processo de ensino-aprendizagem.

Referências

- AVALIA. *AvaliA*. 2019. <<https://siteavalia.grupoa.com.br>>, visitado em 16/09/2022. Citado na página 12.
- BONTHU, S.; SREE, S. R.; PRASAD, M. K. Automated short answer grading using deep learning: A survey. In: SPRINGER. *International Cross-Domain Conference for Machine Learning and Knowledge Extraction*. [S.l.], 2021. p. 61–78. Citado 2 vezes nas páginas 20 e 21.
- BOUD, D.; MOLLOY, E. Rethinking models of feedback for learning: the challenge of design. *Assessment & Evaluation in higher education*, Taylor & Francis, v. 38, n. 6, p. 698–712, 2013. Citado na página 10.
- CAVALCANTI, A. P. et al. Automatic feedback in online learning environments: A systematic literature review. *Computers and Education: Artificial Intelligence*, Elsevier, v. 2, p. 100027, 2021. Citado 2 vezes nas páginas 10 e 11.
- CUTRONE, L. A.; CHANG, M. Automarking: automatic assessment of open questions. In: IEEE. *2010 10th IEEE International Conference on Advanced Learning Technologies*. [S.l.], 2010. p. 143–147. Citado na página 21.
- DEVLIN, J. et al. BERT: Pre-training of deep bidirectional transformers for language understanding. In: *Proceedings of NAACL-HLT*. [S.l.: s.n.], 2019. p. 4171–4186. doi: 10.18653/v1/N19-1423. Citado na página 20.
- FREEMAN, R.; LEWIS, R. *Planning and implementing assessment*. [S.l.]: Routledge, 2016. Citado na página 10.
- GRIER, R. A. et al. The system usability scale: Beyond standard usability testing. In: SAGE PUBLICATIONS SAGE CA: LOS ANGELES, CA. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*. [S.l.], 2013. v. 57, p. 187–191. Citado na página 24.
- HIGGINS, R.; HARTLEY, P.; SKELTON, A. Getting the message across: the problem of communicating assessment feedback. *Teaching in higher education*, Taylor & Francis Group, v. 6, n. 2, p. 269–274, 2001. Citado na página 10.
- IVANIC, R.; CLARK, R.; RIMMERSHAW, R. What am i supposed to make of this?: the messages conveyed to students by tutors' written comments. Open University Press, 2000. Citado na página 10.
- MANNING, C.; SCHUTZE, H. *Foundations of statistical natural language processing*. [S.l.]: MIT press, 1999. Citado 2 vezes nas páginas 20 e 22.
- MARIN, D. R. P. Automatic evaluation of users' short essays by using statistical and shallow natural language processing techniques. *Advanced Studies Diploma Work, University of Madrid*, Citeseer, 2004. Citado na página 21.

NICOL, D. J.; MACFARLANE-DICK, D. Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. *Studies in higher education*, Taylor & Francis, v. 31, n. 2, p. 199–218, 2006. Citado 2 vezes nas páginas 10 e 13.

PARDO, A. et al. Ontask: Delivering data-informed, personalized learning support actions. *Journal of Learning Analytics*, v. 5, n. 3, p. 235–249, 2018. Citado na página 12.

SIDDIQI, R.; HARRISON, C. J.; SIDDIQI, R. Improving teaching and learning through automated short-answer marking. *IEEE Transactions on Learning Technologies*, IEEE, v. 3, n. 3, p. 237–249, 2010. Citado na página 21.

SINGH, A. et al. Gradescope: a fast, flexible, and fair system for scalable assessment of handwritten work. In: *Proceedings of the fourth (2017) acm conference on learning@scale*. [S.l.: s.n.], 2017. p. 81–88. Citado na página 12.

TSAI, Y.-S. et al. Student appreciation of data-driven feedback: A pilot study on ontask. In: *LAK21: 11th International Learning Analytics and Knowledge Conference*. [S.l.: s.n.], 2021. p. 511–517. Citado na página 12.

WANG, Y.; QIN, J.; WANG, W. Efficient approximate entity matching using jaro-winkler distance. In: SPRINGER. *International Conference on Web Information Systems Engineering*. [S.l.], 2017. p. 231–239. Citado na página 21.

WIGGINS, G. *Educative Assessment. Designing Assessments To Inform and Improve Student Performance*. [S.l.]: ERIC, 1998. Citado na página 10.

YUJIAN, L.; BO, L. A normalized levenshtein distance metric. *IEEE transactions on pattern analysis and machine intelligence*, IEEE, v. 29, n. 6, p. 1091–1095, 2007. Citado na página 21.