Análise da aceitação de recomendações explicadas de recursos educacionais para apoiar o ensino e a aprendizagem em um ambiente educacional online.

Randerson Douglas Ribeiro¹, Ig Ibert Bittencourt¹, Ranilson Oscar Paiva¹

¹Instituto de Computação – Universidade Federal de Alagoas (UFAL)

Av. Lourival Melo Mota, S/N - Tabuleiro do Martins

Maceió AL - CEP 57.092-900 – Maceió – AL – Brazil

{rdrs, ig.ibert, ranilsonpaiva}@ic.ufal.br

Abstract. *This article aims to assess the impact of the explanation factor on the acceptance of a recommendation. In order to carry out the analysis, data from the students' interaction with the pedagogical resources available in an online course were simulated. These data were used to measure student performance and interaction. For each scenario obtained, recommendations were made that offered resources based on the student's particular need. Half of these recommendations were explained and the other half were not. In the end, participants were asked whether they would accept such recommendations and whether they were appropriate. And we concluded that the participants had a better perception and were more likely to accept the recommendations explained.*

Resumo. *Este artigo visa avaliar o impacto do fator explicação na aceitação de uma recomendação. Para efetuar a análise, foram simulados dados da interação dos alunos com os recursos pedagógicos disponíveis em um curso online. Esses dados serviram para medir o desempenho e a interação do aluno. Para cada cenário obtido, foram feitas recomendações que ofereciam recursos com base na necessidade particular do aluno. Metade dessas recomendações foram explicadas e a outra não. Ao final, os participantes foram questionados se aceitariam tais recomendações e se elas foram apropriadas. E concluímos que os participantes tiveram uma melhor percepção e foram mais propensos a aceitar as recomendações explicadas.*

1. Introdução

Os ambientes online de aprendizagem, visam oferecer aos alunos uma forma diferente da convencional, no quesito ensino-aprendizagem. Para isso são oferecidos diversos recursos, como: vídeos, questões, textos, chats, links externos e outros. O objetivo é ajudar na absorção e fixação do conteúdo [Paiva and Bittencourt 2017]. Além da diversidade de recursos presentes nesses ambientes, pode-se observar uma variedade quanto à complexidade presente em questões de um mesmo assunto. E isso faz com que o estudante interaja, apenas, com parte desses recursos. Assim cabe aos professores e/ou sistema, ou até mesmo aos próprios estudantes, a tarefa de escolher os melhores recursos, visando a aprendizagem de um determinado assunto [Paiva and Bittencourt 2020]. Esse processo de escolha se repete para cada assunto diferente que é lecionado por meio da plataforma e é possível percebermos isso como um desafio. Sendo assim, alguns problemas podem

surgir, pois diante de toda essa gama de recursos, o aluno pode se sentir confuso, não sabendo com quais recursos interagir para obter o melhor desempenho possível, e segundo [Borg 2017] esse fenômeno é chamado de paralisia por análise, tendo em vista que o excesso de recursos/fatos/informações podem acarretar, mentalmente, a uma suspensão da decisão.

Os sistemas de recomendação (SR) são uma possível solução para esse problema. Sua finalidade é ajudar os usuários a tomarem decisões com base em diversos aspectos dos itens (recursos educacionais) e/ou do usuário (modelo do estudante). Ainda assim, o usuário pode não sentir confiança nas recomendações recebidas, discordando que se aplique ao seu perfil ou não entender o motivo de tê-la recebido, independente da qualidade da recomendação [Abdollahi and Nasraoui 2018]. Contudo, de acordo com [Gunning 2017], a eficácia dos sistemas computacionais é limitada pela incapacidade de explicarem suas decisões e ações (falta de transparência) aos usuários. Diante dessa afirmação, percebemos a importância da transparência no processo de recomendação. De acordo com [Gedikli et al. 2014], a transparência ajuda o usuário a entender a saída e lógica do sistema, melhora a aceitação geral, qualidade percebida e eficácia do sistema. A falta de transparência em sistemas de IA tem se tornado uma preocupação mundial, e foi discutida por representantes de todos os continentes em um evento¹ da UNESCO² ocorrido em 2019. Sendo assim, identificamos a necessidade de explicar ao estudante o motivo de receberem recomendações de itens, tornando tais recomendações mais transparentes, tendo em vista que, as recomendações podem afetar na tomada de decisão, impactando em diversos aspectos da vida dos seres humanos, individualmente ou coletivamente. Tal impacto pode ser injusto se baseado em uma inferência tendenciosa. Para evitar isto, a explicabilidade pode desempenhar um papel importante, promovendo a confiança [Abdollahi and Nasraoui 2018]. E conforme [Tintarev and Masthoff 2012], que em seus estudos têm abordado os objetivos que a explicação proporciona, a autora afirma que, para que uma explicação seja eficaz, ela deve no mínimo obter algum tipo de classificação (uma que reflita as preferências do usuário). Uma explicação que não pode ajudar a obter qualquer classificação, por definição, leva a uma eficácia insuficiente e, além disso, provavelmente resulta em insatisfação do usuário onde sistema provavelmente perderá o usuário. Em segundo lugar, a validade da métrica usada depende do conjunto de dados subjacente.

Dessa forma, identificamos nosso problema de negócio, e o definimos da seguinte forma: Como aumentar a aceitação da recomendação de recursos educacionais? E nosso problema técnico é: Como explicar a recomendação de recursos educacionais em ambientes online de aprendizagem? Propomos, então, a adição de explicações às recomendações de recursos educacionais aos alunos de um ambiente online de aprendizagem, e avaliar os efeitos na aceitação dessas recomendações. Dessa forma, os nossos objetivos são:

- Criar um modelo de cada estudante com base em suas interações com o ambiente online de aprendizagem;
- Criar recomendações automáticas e personalizadas com base no modelo do estudante;
- Mensurar o desempenho pedagógico e de interação do estudante;

¹Evento Mobile Learning Week 2019 - UNESCO

²United Nations Educational, Scientific and Cultural Organization

- Oferecer recomendações, levando em consideração o desempenho do estudante;
- Comparar os cenários de recomendação (explicados e não explicados), avaliando o efeito da explicação nas recomendações quanto à aceitação.

Dessa forma, temos a seguinte questão de pesquisa e suas hipóteses:

QP - Qual o efeito da explicação de recomendações pedagógicas na aceitação dessas recomendações por parte dos estudantes de ambientes online de aprendizagem?

- H_0 - Não Existe diferença significativa na aceitação das recomendações com explicação e sem explicação.
ACEITAÇÃO (Explicação) = ACEITAÇÃO (Sem Explicação)
- H_A - Existe diferença significativa na aceitação das recomendações com explicação e sem explicação.
ACEITAÇÃO (Explicação) $\neq$ ACEITAÇÃO (Sem Explicação)

2. Proposta

A proposta deste trabalho visa criar recomendações explicadas e personalizadas, de forma automática, com base no modelo do estudante (considerando o desempenho do estudante na resolução de questões, o que denominamos: desempenho pedagógico; e nas suas interações, que chamamos de: desempenho de interação dentro do ambiente de aprendizagem online). E a partir desses dados, comparar a aceitação das recomendações pelos estudantes, quando essas estão, acompanhadas ou não, de uma explicação. Foi desenvolvido um experimento onde simulamos as interações de 30 estudantes (seguindo uma distribuição normal) com um curso de estatística básica online. Foi escolhido um caso de um estudante, aleatoriamente, para criar recomendações personalizadas com e sem explicação.

Com isso, o modelo de cada estudante é constituído pelo cálculo do desempenho e da interação para cada tópico que o estudante interagiu. O modelo do estudante definirá o conhecimento do sistema sobre suas respectivas necessidades. No contexto desta pesquisa, definimos como 60%³ o desempenho mínimo a ser alcançado em cada assunto estudado. Não atingir esse percentual para o desempenho pedagógico ou para o desempenho de interações, em um determinado assunto, foi interpretado como dificuldade do estudante no assunto, fazendo com que o sistema gerasse recomendações para ajudá-lo a superar essa dificuldade.

Referente ao desempenho de interação, as interações realizadas pelos estudantes com os recursos educacionais oferecidos pelo sistema foram observadas. O sistema, de forma automática, contabilizou os recursos que os estudante interagiu (vídeoaulas, textos, *links* para recursos em sites externos. Para as questões, também foram obtidas suas dificuldades, se o estudante a respondeu corretamente ou não e o tempo de resposta.

Com esses dados, foi possível identificar os recursos que não foram acessados e inferir o conhecimento do estudante sobre um determinado assunto, e assim, criarmos recomendações que atendam às necessidades desse estudante (como uma espécie de ferramenta de autoria) [Dermeval et al. 2018]. Referente ao desempenho pedagógico, para

³Este valor foi escolhido para atender as demandas do curso descrito no texto, adotado simplesmente por ser um valor que corresponde mais da metade do que foi proposto. Mas pode ser modificado para atender diferentes expectativas/demandas.

os estudantes que obtiveram um desempenho inferior ao valor de referência (60% para esta pesquisa), o sistema recomenda questões referentes ao(s) assunto(s) onde o baixo desempenho foi identificado.

Mediante a análise da performance do estudante, serão oferecidas recomendações acompanhadas de uma explicação que detalha o motivo de tal recomendação ter sido realizada. Os elementos considerados necessários para criação das recomendações foram:

- **Modelo do estudante:** dados simulados que foram utilizados para calcular o desempenho de interação e pedagógicos dos estudantes para cada tópico do curso considerado. Através desse modelo, identificamos os pontos fortes e fracos de cada estudante. O modelo poderia ser atualizado sempre que o estudante realizasse novas interações.
- **Filtros:** os filtros são um conjunto de consultas pré-definidas das interações dos estudantes com os recursos de aprendizagem disponíveis e com o sistema (interações realizadas e as não realizadas).
- **Recursos de Aprendizagem:** são recursos (vídeoaulas, textos, questões, apresentações de aula e links para recursos de outros sites) associados a um tópico do curso, e que são disponibilizados pelo sistema com o objetivo de promover a aprendizagem dos estudantes.
- **A Recomendação dos Recursos de Aprendizagem:** as recomendações dos recursos ocorreram por meio de notificações *push* que o ambiente de aprendizagem enviava para os estudantes alvo. As recomendações eram exibidas no canto superior direito da tela, podendo o estudante aceitá-la ou recusá-la.
- **A Explicação:** é a parte do texto que justifica/explica a recomendação, tornando-a transparente para o usuário, ou seja, é o que justifica o motivo pelo qual a recomendação foi enviada ao estudante.

3. Métodos

Para a validação da hipótese desta pesquisa, foi enviado um questionário para 300 estudantes dos cursos de graduação em ciência da computação e engenharia da computação.

Em resumo, utilizamos um ambiente online de aprendizagem, e nele criamos um curso de estatística descritiva. Em decorrência de algumas instabilidades do ambiente de aprendizagem, decidimos simular (seguindo uma distribuição estatística normal) as interações dos estudantes com os recursos educacionais disponibilizados. Os dados de um (o estudante com ID 8) desses 30 estudantes simulados (*personas*), foram escolhidos aleatoriamente para criar 4 (quatro) cenários de recomendação (2 cenários com a recomendação explicada e outros 2 com recomendação sem explicação) e para calculamos o desempenho do estudante (na resolução de problemas e referente às interações).

A organização e fluxo do experimento funcionaram da seguinte forma: os participantes, após aceitarem participar do experimento, deveriam ler um resumo da situação pedagógica do estudante 8. Em seguida, deveriam ler a recomendação pedagógica criada para esse estudante, e definir o seu nível de concordância em uma escala *Likert*, subseção 4.2).

4. Resultados e Discussões

Neste capítulo reportaremos o resultado do questionário e, em seguida, discutiremos esses resultados.

4.1. Participantes

O questionário ficou disponível do dia 22/10/2019 até o dia 11/11/2019. Obtivemos 32 participações válidas, compostas por professores, graduandos e graduados dos cursos do Instituto de Computação (em ciência da computação, engenharia da computação), assim como mestres e mestrandos em Modelagem Computacional e Mestrado em Informática.

Dentre os participantes, 70% eram do sexo masculino e 30% do sexo feminino. Em relação à ocupação dos participantes, 83% dos participantes preencheram esta informação: 73% são estudantes; 8% são professores; e 12% exercem outras atividades. Em relação ao nível de escolaridade: 50% dos participantes estão cursando a graduação; 20% possuem a graduação completa; 16,7% possuem especialização; 10% possuem mestrado e 3,3% possuem doutorado. Em relação ao nível de conhecimento no uso de tecnologias digitais: 60% informaram ter um conhecimento avançado; 36,7% nível moderado e 3,3% nível baixo. Em relação à participação em cursos online: 96,3% dos participantes já participou de algum curso online e apenas 3,7% nunca participou de nenhum curso online.

Percebemos que o número de elementos da amostra é limitado, mas foi uma quantidade que nos permitiu realizar algumas análises e inferências.

4.2. Respostas ao Questionário

Conforme informado, foram elaborados 4 cenários de recomendação (descritos adiante nesta subseção), com base no modelo do estudante com ID 8, conforme descrito abaixo.

O aluno 8, obteve um baixo percentual de desempenho pedagógico (PDP), no assunto VARIÂNCIA E DESVIO PADRÃO (assunto 3), do módulo MEDIDAS DE DISPERSÃO (módulo 3). Observou-se que das 4 questões que foram oferecidas pelo sistema, o aluno respondeu 3, acertando 1 e errando 2, obtendo um PDP de 25%. O desempenho aceitável pelo sistema é de, no mínimo, 60%.

A Figura 1 apresenta as respostas dos participantes para cada uma das 6 afirmações apresentadas após cada uma das 4 recomendações criadas, para abordar a situação do estudante 8. Os participantes manifestaram o nível de concordância para cada afirmação, escolhendo uma opção de uma escala *Likert* com 5 alternativas: Discordo Completamente (DC), Discordo (D), Não Concordo nem Discordo (NCND), Concordo (C) e Concordo Completamente (CC). As afirmativas foram:

AF 01	A recomendação oferecida foi apropriada (justifica o problema descrito)
AF 02	O texto da recomendação explica o motivo pelo qual a recomendação foi enviada.
AF 03	O texto da recomendação faria (motiva) que a recomendação seja seguida.
AF 04	A recomendação está clara (sou capaz de entendê-la e saber o que devo fazer).
AF 05	Se eu fosse o aluno 8, eu seguiria essa recomendação.
AF 06	Acho importante que o sistema explique porque uma recomendação foi gerada.

Tabela 1 - Afirmações apresentadas aos participantes.

As recomendações foram ofertadas de forma alternada, onde a primeira recomendação foi entregue sem explicação, a segunda foi acompanhada de uma explicação, em seguida a terceira sem explicação e a última com explicação.

1. Sem Explicação “Olá estudante 8, tudo bem? Sugerimos que você reveja a apresentação da aula e responda as questões sobre VARIÂNCIA E DESVIO PADRÃO (contidas no módulo MEDIDAS DE DISPERSÃO).”
2. Com Explicação “Olá estudante 8, tudo bem? Observamos que seu desempenho não foi bom o suficiente no assunto VARIÂNCIA E DESVIO PADRÃO, desse modo, pedimos que por gentileza, assista ao vídeo sobre VARIÂNCIA E DESVIO PADRÃO (contido no módulo MEDIDAS DE DISPERSÃO) e acesse o link externo do assunto, como forma de revisão, para que possa melhorar seu desempenho na resolução das questões (tendo em vista que seu coeficiente de desempenho pedagógico foi apenas 25%). Observou-se que das 4 questões oferecidas, 3 foram respondidas, mas somente uma de forma correta!”
3. Sem Explicação “Olá estudante 8, tudo bem? Pedimos por gentileza que leia o texto auxiliar, acesse o link externo, assista novamente ao vídeo sobre o assunto VARIÂNCIA E DESVIO PADRÃO (contido no módulo MEDIDAS DE DISPERSÃO) e responda as questões desse assunto.”
4. Com explicação “Olá estudante 8, tudo bem? Observamos que seu desempenho de interação foi satisfatório (87,5%) e o parabenizamos por isso, pois vimos que utilizou quase todos os recursos disponíveis no sistema, mas percebemos que o seu desempenho pedagógico (resolução de questões) não foi bom (25%), para o assunto VARIÂNCIA E DESVIO PADRÃO. Por isso, pedimos por gentileza, que assista novamente o vídeo sobre VARIÂNCIA E DESVIO PADRÃO (contido no módulo MEDIDAS DE DISPERSÃO) e reveja a apresentação da aula que contém exemplos bastante didáticos, que te ajudarão a resolver todas as questões.”

Afirmações	Cenário 1					Cenário 2					Cenário 3					Cenário 4				
	DC	D	NCND	C	CC	DC	D	NCND	C	CC	DC	D	NCND	C	CC	DC	D	NCND	C	CC
AF 01	0	16.7	10	60	13.3	0	13.3	3.3	53.3	30	6.7	30	13.3	40	10	0	3.3	0	50	46.7
AF 02	23.3	46.3	10	20	0	0	0	0	60	40	26.7	46.7	10	16.7	0	0	3.3	0	26.7	70
AF 03	6.7	23.3	36.7	30	3.3	3.3	16.7	10	43.3	26.7	16.7	23.3	30	23.3	6.7	6.7	3.3	6.7	30	53.3
AF 04	0	0	3.3	63.3	33.3	0	3.3	0	60	36.7	3.3	6.7	23.3	50	16.7	0	0	3.3	30	66.7
AF 05	0	6.7	16.7	66.7	10	0	10	10	43.3	36.7	3.3	23.3	23.3	43.3	6.7	0	3.3	10	30	56.7
AF 06	0	0	33.3	30	66.7	0	0	0	36.7	63.3	0	0	7	26.3	66.7	0	0	0	16.7	83.3

Figura 1. Resultados em valores percentuais.

4.3. Discussão dos Resultados

De forma geral, percebemos que as recomendações explicadas (cenários 2 e 4) obtiveram um melhor resultado que as recomendações não explicadas (cenários 1 e 3), em relação à percepção dos participantes, para todos os critérios avaliados. Isso sugere que a explicação favorece a aceitação das recomendações. Abaixo, descreveremos os resultados para cada afirmação contida no questionários, para cada um dos cenários de recomendação.

Utilizamos o método de Wilcoxon para comparar as performances de cada cenário (pares de cenários) no sentido de verificar se existem diferenças significativas entre os seus resultados, que são exibidos nas tabelas 2, 3, 4, 5, 6 e 7.

Afirmiação 01	Cenário 1	Cenário 2	Cenário 3	Cenário 4
Cenário 01	-	0,1946	0.009032	0.00767
Cenário 02	-	-	0.01538	0.1622
Cenário 03	-	-	-	0.0002763

Tabela 2 - Wilcoxon comparação da AF 01 de todos os cenários

Em relação a primeira afirmação (AF 01), para o cenário 1, percebemos que houve uma quantidade considerável de pessoas que concordam que a recomendação oferecida foi apropriada (apesar de não ser explicada), mas vale ressaltar que a finalidade era que todas as recomendações tivessem qualidade similar e satisfatória, de forma a não influenciar no resultado. No cenário 2 (com explicação), percebemos um aumento na concordância, que volta a reduzir no cenário 3 (sem explicação) e, novamente, aumenta no cenário 4 (com explicação). Isso nos leva a crer que a variação da percepção para AF 01 se deve, à falta da explicação. Nos casos 1 e 4, 2 e 3, 3 e 4 ocorreu o que esperávamos (cenários diferentes) houve relevância. Observamos que não houve diferença significativa entre os cenários 2 e 4 (cenários com explicação).

Afirmação 02	Cenário 1	Cenário 2	Cenário 3	Cenário 4
Cenário 01	-	1.945e-06	0.766	3.558e-06
Cenário 02	-	-	3.313e-06	0.06371
Cenário 03	-	-	-	2.073e-06

Tabela 3 - Wilcoxon comparação da AF 02 de todos os cenários

Sobre a segunda afirmação (AF 02), o cenário 1 (não explicado), apresenta baixo percentual de concordância, apenas 20%. A recomendação não era explicada e os participantes conseguiram perceber que o texto não estava convincente e faltava algo. Já na segunda recomendação (explicada) obtivemos 100% de concordância, de que o texto fornece elementos que tornam possível a compreensão de tal oferta. Isso nos fornece fortes indícios de que a criação do texto da recomendação foi apropriada e que a explicação cumpriu sua função. Com a terceira recomendação percebemos, novamente, um alto percentual de discordância (73,4%). Este resultado indica que os participantes não identificaram uma explicação na recomendação, reforçando o fato de que eles compreenderam o conceito de recomendação explicada. A quarta recomendação traz um altíssimo percentual de concordância (96,7% concordam). Diante do teste de Wilcoxon vimos que, ao comparar a mesma afirmação em cenários distintos (explicado X não explicados), houve uma diferença estatisticamente relevante, onde o p-valor foi $< 0,05$. Observamos que os estudantes conseguiram perceber a presença da explicação e diferenciá-la de uma recomendação não explicada por meio do texto apresentado.

Afirmação 03	Cenário 1	Cenário 2	Cenário 3	Cenário 4
Cenário 01	-	0.001575	0.5609	0.0004368
Cenário 02	-	-	0.01121	0.008123
Cenário 03	-	-	-	0.0008265

Tabela 4 - Wilcoxon comparação da AF 03 de todos os cenários

Quanto a terceira afirmação (AF 03), a primeira recomendação (sem explicação) mostra que apenas 33,3% dos participantes concordam, que o texto da recomendação motiva sua aceitação. Este resultado apresenta indícios da importância da explicação das recomendações para motivar sua aceitação. Na segunda recomendação (explicada), percebemos um alto percentual de concordância (70%). Comparando com o resultado para a recomendação 1, percebemos indícios de que a explicação motiva a aceitação da recomendação. Na terceira recomendação, percebemos uma diminuição na concordância

que cai para 30% reforçando que a explicação afeta a aceitação da recomendação. A quarta recomendação mostra, um alto percentual de concordância (83,3%). O resultado é o maior dentre as recomendações apresentadas aos participantes neste cenário, sugerindo que a explicação é um fator importante para a aceitação da recomendação. Para o teste de Wilcoxon, ao comparar os resultados de todos os cenários, vimos que a comparação de cenários diferentes mostrou $p\text{-valor} < 0.05$, ou seja, houve significância estatística, concordando com a nossa expectativa, que seria a percepção do usuário em relação ao texto, diferenciando-os. Já na comparação de cenários iguais, o resultado dos cenários não explicados foram condizentes com a nossa expectativa, ou seja, que não houvesse diferença estatística significativa ($p\text{-valor} > 0.05$). Mas nos cruzamento dos cenários explicados (cenários 2 e 4), vimos que houve uma diferença significativa, onde acreditávamos que não haveria. Diante disso, acreditamos que os participantes devem ter percebido que a recomendação do cenário 4 foi melhor elaborada, pois continha um nível maior de explicação e detalhamento em relação a recomendação 2.

Afirmção 04	Cenário 1	Cenário 2	Cenário 3	Cenário 4
Cenário 01	-	0.8353	0.001443	0.01383
Cenário 02	-	-	0.0128	0.008457
Cenário 03	-	-	-	0.0001645

Tabela 5 - Wilcoxon comparação da AF 04 de todos os cenários

A quarta afirmação (AF 04), traz um percentual de 96% de concordância e o percentual de discordância foi 0. isso mostra que apesar da recomendação inicial oferecida ser sem explicação, os participantes foram capazes de compreender quanto ao que deveriam fazer. Na segunda recomendação, o índice de concordância se manteve alto (97,6%). Dessa forma não houve diferença significativa, com relação a percepção da primeira recomendação, e isso gera indícios de que a explicação pode não afetar a clareza da recomendação como um todo. Sendo assim, como a explicação é parte integrante da recomendação, se a recomendação estiver clara, é possível que a explicação não agregue valor de forma significativa. Na terceira recomendação (explicada), observou-se que 66,7% concordam com a afirmação. E isso pode ser um ajuste após a interação com uma recomendação explicada, mas ainda assim, a concordância foi de 2/3, e isso é apropriado para tal critério. Na quarta recomendação (explicada), os resultados foram satisfatórios, mostrando que 96,7% concordam com a afirmação, ao comparar este resultado com o cenário anterior, podemos ver que a explicação motiva a aceitação e o engajamento com a recomendação. No teste de Wilcoxon, comparação feita entre os cenários 1 e 4; 2 e 3; 3 e 4, mostra que houve diferença estatisticamente relevante, pois o $p\text{-valor}$ foi $p < 0.05$. O usuário conseguiu perceber a diferença entre eles no decorrer da realização do questionário, o que foi concordante com a nossa expectativa. Já na avaliação dos cenários 2 e 4, pela similaridade contextual, houve uma relevância, o que consideramos normal, pois os participantes devem ter percebido que a recomendação 4 tinha mais detalhes em relação à recomendação 2.

Afirmção 05	Cenário 1	Cenário 2	Cenário 3	Cenário 4
Cenário 01	-	0.08759	0.007762	0.001627
Cenário 02	-	-	0.005874	0.0495
Cenário 03	-	-	-	0.0003657

Tabela 6 - Wilcoxon comparação da AF 05 de todos os cenários

Na quinta afirmativa (AF 05), a primeira recomendação trouxe um alto percentual de concordância (76,7%), com esse indicador pode-se que a recomendação está apropriada. A segunda recomendação, houve um leve aumento na concordância (80%), e isso sugere que a qualidade da recomendação (independente da explicação) está apropriada. No terceiro cenário (não explicado), o resultado foi menor (50%) em relação aos cenários 1 e 2, evidenciando que a falta de explicação influencia na decisão dos participantes (se seguem ou não a recomendação). A quarta recomendação (explicada), conforme esperado a presença da explicação fez com que houvesse um aumento significativo na aceitação e engajamento com a recomendação, atingindo 86,7%. Com o teste de Wilcoxon, vimos que, nas comparações feitas nos cenários: 1 e 4, 2 e 3, 3 e 4 foram significativamente relevantes pois o p-valor obtido foi $p < 0.05$ e isso coincidiu com a nossa expectativa, que versava haver diferença significativa na comparação entre cenários diferentes, onde os cenários com explicação obteriam um melhor desempenho em relação aos cenários sem explicação.

Afirmiação 06	Cenário 1	Cenário 2	Cenário 3	Cenário 4
Cenário 01	-	1	0.8211	0.09534
Cenário 02	-	-	0.8241	0.04108
Cenário 03	-	-	-	0.02627

Tabela 7 - Wilcoxon comparação da AF 06 de todos os cenários

Diante da sexta afirmativa (AF 06), os resultados foram significativos com altos índices percentuais em todos os cenários, onde os participantes consideraram que, é importante que um sistema explique o motivo de uma recomendação ter sido gerada. Onde o menor índice percentual foi de 93%. Diante do teste de Wilcoxon, nossa expectativa era que independente do tipo de cenário (explicado ou não explicado), não haveria diferença estatística significativa, onde era esperado um p-valor > 0.05 . E houve uma concordância, observando os resultados obtidos, onde não houve diferença estatística significativa na maioria dos casos. Em apenas dois casos, os resultados foram significantes com p-valor < 0.05 , que foram as comparações dos cenário 2 e 4; e 3 e 4, onde no primeiro caso (2 e 4) pelo fato de ambas as recomendações apresentarem explicações o usuário achou que já estava suficiente o bastante, mas não descartando a sua importância. No segundo caso (3 e 4), acreditamos que a significância foi gerada justamente pelo fato do usuário entender melhor a recomendação 4 (explicado) devido ao maior nível de detalhe em relação da recomendação da recomendação 3 (não explicado), onde nesse caso o usuário pode ter considerado que a afirmativa se referia a questão vigente e não ao contexto geral, o qual de fato era a situação avaliada.

5. Conclusão, Limitações e Trabalhos futuros

Observando os resultados alcançados, podemos concluir que no contexto geral, os participantes consideraram que as recomendações foram apropriadas e justificam a recomendação, tendo uma melhor percepção à partir do 2º cenário (explicado). Pode-se dizer também, que o texto da recomendação explica o motivo pelo qual ela foi enviada, apesar de não haver muita diferença comparado ao 1º cenário, mas percebemos melhores índices quando o cenário é explicado, evidenciando ainda mais que a explicação é um fator que favorece a tomada de decisão do usuário. Os participantes consideram que o texto da recomendação (motiva) o seguimento (adesão) da mesma, onde diante dos indicadores, é notório que a explicação afeta a aceitação da recomendação. Quanto à clareza da

recomendação, observamos que se a recomendação está clara, o fator explicação pode passar despercebido e não agregar valor de forma tão significativa. Quanto ao fato de seguir ou não a recomendação, os dados apontam que os cenários explicados fazem com que haja uma maior pretensão dos usuários em seguir a recomendação. De modo geral, a presença da explicação na recomendação é importante, pois ela permite que o usuário possam discernir e tomar decisões de forma consciente.

Quanto às limitações, é que os dados utilizados foram simulados e podem não representar completamente a situação de um curso com alunos reais, mas foi uma forma plausível de simular os dados, seguindo uma distribuição normal que alguns cursos não seguem, sendo eles ou muito fáceis ou muito difíceis enviesando para um lado ou outro. A quantidade de participantes não foi o ideal, mas foi uma boa quantidade de participantes, mas isso pode limitar a capacidade de generalização dos resultados. A análise da aderência também passou a ser uma limitação, pois o termo aderência está ligado ao fato de aceitar e fazer, onde só foi possível analisar a aceitação, pois o ato de fazer está diretamente ligado a ações das pessoas.

Diante destas informações, pretendemos utilizar um ambiente de um curso com alunos reais, abrangendo um tempo maior, testar diferentes tipos de recomendações enviadas para os alunos em para diferentes disciplinas, e não somente estatística. Aprimorar o modelo do aluno, para que ele possa ser um modelo aberto de modo que o usuário possa afetar explicitamente o próprio modelo, podendo ser através de opiniões que possibilitem a validação quanto à assertividade da recomendação para a necessidade observada.

Referências

- Abdollahi, B. and Nasraoui, O. (2018). Transparency in fair machine learning: the case of explainable recommender systems. In *Human and Machine Learning*, pages 21–35. Springer.
- Borg, J. (2017). *A arte da persuasão*. Saraiva Educação SA.
- Dermeval, D., Paiva, R., Bittencourt, I. I., Vassileva, J., and Borges, D. (2018). Authoring tools for designing intelligent tutoring systems: a systematic review of the literature. *International Journal of Artificial Intelligence in Education*, 28(3):336–384.
- Gedikli, F., Jannach, D., and Ge, M. (2014). How should i explain? a comparison of different explanation types for recommender systems. *International Journal of Human-Computer Studies*, 72(4):367–382.
- Gunning, D. (2017). Explainable artificial intelligence (xai). *Defense Advanced Research Projects Agency (DARPA)*, nd Web.
- Paiva, R. and Bittencourt, I. I. (2017). Helping mooc teachers do their job. In *Researcher Links Workshop: Higher Education for All*, pages 52–67. Springer.
- Paiva, R. and Bittencourt, I. I. (2020). Helping teachers help their students: A human-ai hybrid approach. In *International Conference on Artificial Intelligence in Education*, pages 448–459. Springer.
- Tintarev, N. and Masthoff, J. (2012). Evaluating the effectiveness of explanations for recommender systems. *User Modeling and User-Adapted Interaction*, 22(4-5):399–439.