

AvaliaSmart: Avaliação Automática de Questões Discursivas

Lucas Busatta Galhardi¹, João Antônio Mendes Oliveira¹, Caio Augusto G. Cavazzani¹, Rodrigo Clemente Thom de Souza^{2,3,4}

¹AvaliaSmart – Correções Automáticas de Questões Discursivas
Hotel Tecnológico da Universidade Tecnológica Federal do Paraná
86.001-970 - Londrina - PR - Brazil

²Advanced Campus in Jandaia do Sul
Federal University of Paraná (UFPR) 86900-000 - Jandaia do Sul, PR – Brazil

³Graduate Program in Computer Science
State University of Maringá (UEM) 87020-900 Maringá, PR - Brazil

⁴Graduate Program in Production Engineering
State University of Maringá (UEM) 87020-900 Maringá, PR - Brazil
{ti,gerencial,comercial}@avaliasmart.com.br, thom@ufpr.br

Resumo. Este artigo apresenta brevemente a AvaliaSmart, startup que possui como principal aplicação um modelo para avaliação automática de questões discursivas. Esse tipo de avaliação automática é inédita no Brasil e possui como principal benefício o auxílio a alunos e professores, podendo interessar à diversas instituições ligadas à área da educação que queiram otimizar esse processo internamente. O artigo descreve como a ferramenta passou pelas etapas de identificação do problema, ideia, pesquisa científica até chegar ao mercado como um produto/serviço.

1. Ideia

Uma das etapas importantes no processo de aprendizado é a avaliação do aluno. Essa avaliação pode consistir em provas, trabalhos e atividades diversas. Dentre as possibilidades, o professor pode optar por utilizar questões que o aluno deve responder de alguma forma, sendo as mais conhecidas as de múltipla escolha e as discursivas. Enquanto a primeira já se beneficia de uma correção automatizada por computadores há muitos anos, as questões dissertativas não possuem uma solução definitiva até hoje. Essa facilidade na correção leva muitos professores a optarem pelas questões de múltipla escolha. Porém, esse tipo de questão pode não avaliar o aluno tão profundamente quanto questões discursivas, que exigem uma resposta totalmente elaborada pelo aluno no momento da atividade [Burrows 2015].

Entretanto, ao se utilizar as questões discursivas na prática, outros problemas podem surgir. Salas de aulas geralmente possuem dezenas de alunos, quando não as centenas praticadas por cursos em formato EAD. Se o professor aplica algumas questões discursivas logo ele tem em suas mãos o trabalho de avaliar, uma a uma, dezenas e até centenas de respostas. Isso acarreta professores levando trabalho para casa

[Jacomini and Penna 2016] e comprometendo seu tempo que poderia ser utilizado em outras atividades mais relevantes [Nascimento and Santos 2015].

Além disso, o modelo de correção manual por professores e tutores sofre do problema de falta de padronização: dependendo de variáveis humanas como o humor, cansaço e viés (de respostas anteriores) os avaliadores podem avaliar diferentemente respostas iguais ou muito parecidas ou até mesmo atribuir a mesma nota para respostas bem diferentes em seu nível de precisão [Burrows 2015]. Por fim, o modelo manual também sofre com sua escalabilidade. Com o crescimento do EAD, Ambientes Virtuais de Aprendizagem e MOOCs [ABED 2016], torna-se inviável o uso de questões discursivas em ambientes com centenas e milhares de alunos.

A partir desse cenário é possível concluir que uma ferramenta que avaliasse de maneira correta, automática, padronizada e rápida poderia possibilitar o uso de mais questões discursivas no ambiente de ensino, melhorando a aprendizagem dos alunos e possibilitando que o tempo do professor possa ser usado em atividade mais relevantes. Apesar das claras vantagens, ainda não existem soluções como essa no mercado brasileiro. Mesmo as pesquisas científicas da área são muito restritas no país (Seção 2).

Esse contexto traz a AvaliaSmart como uma startup com uma proposta inovadora, que oferece a funcionalidade de avaliação automática como um serviço de um sistema de informação, possibilitando a rápida difusão da ferramenta pelo mercado.

2. Pesquisa

A ideia da avaliação automática de questões discursivas curtas foi trabalhada durante todo um projeto de mestrado. Nesse projeto foram realizadas diversas etapas e experimentos, iniciando por uma pesquisa na literatura científica. Uma das primeiras e mais completas revisões da literatura encontradas foi o trabalho de [Burrows 2015] que definiu os termos da área e categorizou os problemas e métodos possíveis até então.

A partir disso, foi-se escolhido um método específico (Aprendizado de Máquina Supervisionado) e feito uma revisão sistemática da literatura. Esse trabalho encontrou 44 publicações em inglês que abordavam o problema utilizando esse método [Galhardi and Brancher 2018]. A partir de então, um novo método foi criado baseado na junção de outros trabalhos da literatura e um experimento comprovou sua eficácia nos conjuntos de dados mais relevantes da área de pesquisa (em inglês) [Galhardi et al. 2018b].

Entretanto, a falta de trabalhos com conjuntos de dados em português levou à uma pesquisa da literatura na língua portuguesa, estendendo a busca a outros métodos que não apenas o aprendizado de máquina. Essa pesquisa encontrou 7 trabalhos brasileiros relacionados, que lidavam com pequenos conjuntos de dados. A quantidade escassa de trabalhos e a inexistência de um conjunto de dados maior na língua portuguesa levou à dúvida sobre a eficácia do modelo para o português também. Portanto, o próximo passo foi criar um conjunto de dados reais, a partir de alunos brasileiros do ensino fundamental e médio, adaptar e melhorar a ferramenta e realizar novos experimentos [Galhardi et al. 2018a].

Por fim, um último trabalho reuniu os últimos avanços da área e realizou novos experimentos de maneira sistemática, produzindo resultados que possam ser comparados com novos trabalhos [Galhardi et al. 2020] e também gerando um registro de software do modelo [Galhardi and Brancher 2019].

3. Engenharia de Construção e Produção

O "produto" final obtido durante o projeto de mestrado foi denominado simplesmente de "algoritmo", consistindo do código de treinamento e o de predição, funcionando para quaisquer questões ao se prover os dados necessários. Para a transformação do algoritmo em um serviço pronto para o mercado foram necessários alguns passos:

- Análise do Mercado: entendimento dos processos atuais do mercado no tratamento desse tipo de dados (questões, respostas e avaliações);
- Projeto Técnico: consultoria com uma empresa de TI com mais experiência no desenvolvimento de sistemas do tipo e criação do projeto técnico do sistema;
- Construção do sistema utilizando um processo ágil, com participação ativa dos stakeholders para refinar os requisitos, funcionalidades e integrações.

A partir desse processo foi-se criada toda a infraestrutura da AvaliaSmart, que inclui:

- Algoritmo encapsulado em um container, com um endpoint de treinamento e outro de predição, com toda a infraestrutura na Nuvem, pronta para ser rodada em escala e com alta disponibilidade (os containers recebem dados de entrada, realizam um processamento, salvam os resultados e são encerrados);
- Painel administrativo para controle de toda a aplicação e processo do ciclo de vida dos dados (há o acesso de usuário diferenciado para sócios e funcionários da startup e o acesso dos clientes). No painel é possível gerenciar: instituições, escalas, questões, modelos treinados, infraestrutura, usuários, integrações, visualizar estatísticas dentre outras funcionalidades complementares;
- Integrações com APIs: via back-end e front-end;
- Demonstrações: páginas especiais que demonstram o funcionamento da AvaliaSmart dentro das instituições parceiras.

4. Comercialização como um novo produto

O serviço da AvaliaSmart está pronto para comercialização há poucos meses, buscando um primeiro parceiro que queira validar a solução em seu ambiente real de utilização. Na página principal do site há maiores detalhes sobre a solução e contato¹. Há também um vídeo explicativo apresentando pontos chave da solução e demonstrando seu funcionamento em tempo real².

O modelo de negócios foi planejado para o público-alvo de instituições de ensino a distância (EAD), pois essas instituições têm sistemas totalmente digitalizados, o que facilitaria a integração entre os sistemas, ao grande volume de alunos e, conseqüentemente, de questões respondidas.

A partir da validação da ideia por meio de pesquisa de mercado, realizou-se o planejamento para a disponibilização do serviço de correção de questões discursivas de maneira automática. A solução encontrada foi o desenvolvimento de uma integração Web Service, por meio de APIs, com o objetivo de interligar os sistemas educacionais

¹ <https://www.avaliasmart.com.br/>

² <https://youtu.be/Bh6fyIK8814>

ao algoritmo de inteligência artificial. Dessa forma, garante-se agilidade, confiabilidade e integridade dos dados.

O modelo de precificação será o Software as a Service (SaaS), com a cobrança sendo realizada de acordo com a demanda de utilização do cliente. O modelo de negócios se apresenta viável para os clientes, prometendo realizar a correção de questões discursivas de uma maneira mais ágil, confiável e com menor custo.

5. Conclusão

Diante de tudo o que foi exposto é possível concluir que a AvaliaSmart possui um sistema de informação funcional e inovador, pronto para tornar a experiência da avaliação de questões discursivas melhor para todos, onde o professor/tutor poderá ter mais tempo para desenvolver um trabalho mais próximo e humano com seus alunos. O sistema já passou por testes e validações técnicas e está pronto para uso. A próxima etapa da startup é a de iniciar testes em um primeiro cliente para coletar mais informações e usar o sistema com uma escala maior de respostas e questões.

Referências

- ABED (2016). Censo EAD Brasil 2016 - Relatório Analítico de Aprendizagem a Distância no Brasil.
- Burrows, S., Gurevych, I., and Stein, B. (2015). The eras and trends of automatic short answer grading. *International Journal of Artificial Intelligence in Education*, pages 60–117.
- Galhardi, L., de Souza, R. C. T., and Brancher, J. (2020). Automatic grading of portuguese short answers using a machine learning approach. In *Anais Estendidos do XVI Simpósio Brasileiro de Sistemas de Informação*, pages 109–124. SBC.
- Galhardi, L. B. and Brancher, J. D. (2019). PSA-AutoGrader. INPI.
- Galhardi, L. B., Barbosa, C. R., de Souza, R. C. T., and Brancher, J. D. (2018a). Portuguese automatic short answer grading. In *Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE)*, volume 29, page 1373.
- Galhardi, L. B. and Brancher, J. D. (2018). Machine learning approach for automatic short answer grading: A systematic review. In *Ibero-american conference on artificial intelligence*, pages 380–391. Springer, Cham.
- Galhardi, L. B., Thom de Souza, R. C., C. M., S. H., and Brancher, J. D. (2018b). Exploring distinct features for automatic short answer grading. In *Encontro Nacional de Inteligência Artificial e Computacional*, page 1. 12.
- Jacomini, M. A. and Penna, M. G. d. O. (2016). Carreira docente e valorização do magistério: condições de trabalho e desenvolvimento profissional. *Proposições*, 27(2):177–202.
- Nascimento, M. d. G. C. d. A. and Santos, J. V. (2015). Sessão Especial 05 - Políticas educacionais e currículo: interfaces na educação infantil e ensino fundamental 37a Reunião Nacional da ANPEd – 04 a 08 de outubro de 2015, UFSC – Florianópolis.