

Universidade Federal do Rio de Janeiro
Instituto de Física
Programa de Ensino de Física

Projeto: Aplicação de Algoritmos de Aprendizagem de Máquina no ensino de Física

Candidato: Dr. Charlie Vargas Sarmiento
Supervisor: Prof. Germano Maioli Penello

Maio - 2022

Aplicação de Algoritmos de Aprendizagem de Máquina no ensino de Física

Resumo

No CEDERJ, modalidade de ensino à distância pública do Estado do Rio de Janeiro, foi percebido que as avaliações tradicionais utilizadas na disciplina de Introdução às Ciências Físicas 1 (ICF1) – denominadas como oficinas apresentavam algumas deficiências considerando seu objetivo principal de auxiliar no aprendizado do aluno. Assim, em 2019, foram implementadas avaliações online na plataforma MODDLE utilizada pelo CEDERJ com diferentes vantagens para os alunos, como por exemplo: a correção da avaliação é imediata e o aluno pode refazer as avaliações o número de vezes necessário até obter a nota desejada. De início, apenas dois tópicos da disciplina foram adaptados para este formato, que obteve uma boa aceitação tanto de alunos como de tutores e professores. Este fato nos incentivou em 2020 a ampliar a iniciativa e implementar as avaliações online para os seis principais tópicos de estudo que compõem a disciplina.

Os baixos níveis de aprovação observados no ensino das ciências puras, principalmente da matemática e da física, indicam um fértil campo de estudos sobre metodologias para o ensino. Com o crescente avanço da tecnologia, muitos métodos e técnicas estão surgindo para produzir uma experiência de qualidade e enriquecer o setor educacional como um todo. Dentre os diferentes métodos, os algoritmos de aprendizagem de máquina (*Machine Learning* – ML – em inglês) estão sendo cada vez mais utilizados porque a ideia principal destes algoritmos é aprender, para isto, se baseiam em experiências passadas (dados coletados). Com isso, a utilização de algoritmos supervisionados é uma excelente alternativa para estudar a evolução do rendimento dos alunos e tentar antecipar as notas dos alunos antes das avaliações somativas.

A grande quantidade de alunos em ICF1 (da ordem de 1500 alunos inscritos por semestre) e a implementação das novas avaliações da disciplina vêm gerando um grande volume de dados. Assim, este projeto tem por objetivo principal utilizar algoritmos de ML para fazer um estudo da evolução das notas obtidas pelos alunos e, a partir de informações adicionais (como município, curso e outros, que serão testadas na implementação dos algoritmos) tentar antecipar as dificuldades encontradas por novos estudantes. A ideia deste projeto é de utilizar algoritmos supervisionados e não supervisionados para criar um sistema que permita identificar antecipadamente alunos em risco de abandono da disciplina e/ou reprovação. Isso permitirá que os coordenadores possam acionar os tutores da disciplina para que façam um contato mais individualizado com o estudante e assim, reduzir o número de reprovações e evasões. Em síntese, esta proposta tem como objetivo fazer com que o ensino de massa, como o utilizado no CEDERJ, possa aproveitar de novas tecnologias para que uma atenção individualizada seja oferecida a estudantes que estejam precisando dessa atenção.

Introdução

O rápido crescimento da tecnologia tem gerado um grande volume de dados, que geralmente ultrapassa as capacidades cognitivas dos seres humanos para analisá-los. Algoritmos de aprendizagem de máquina (ML) estão se transformando em uma ferramenta muito importante para análises de grande volume de dados. A aplicação de algoritmos de ML a grandes bases de dados é conhecida como *Data Mining*, onde se analisa os dados para construir um modelo funcional que ajude a prever futuros resultados. [1]

Os algoritmos de ML, sendo parte do que é conhecido como inteligência artificial, são pensados para aprender a partir das experiências prévias e assim resolver problemas para os quais não foram inicialmente pensados. Esses algoritmos de ML podem ser divididos em quatro categorias, dependendo da sua abordagem.

- A aprendizagem **supervisionada** (*supervised learning*), onde dados categorizados e o resultado final, neste caso poderia ser a situação final do aluno, são entregues ao algoritmo para aprender a partir disso. Com o banco de dados que já possuímos, utilizaremos os dados adquiridos para “treinar” o algoritmo.

- A aprendizagem **não supervisionada** (*unsupervised learning*), onde não é passado nenhum rótulo dos dados e nem a resposta correta. Este enfoque é mais complexo porque o algoritmo precisa determinar os padrões adequados para fazer as previsões.

- A aprendizagem **semi-supervisionada** (*semisupervised learning*) que faz uma mistura das duas anteriores.

- A aprendizagem **reforçada** (*reinforcement learning*), onde o algoritmo recebe recompensas quando segue comportamentos desejados e punições nas indesejadas. [1, 2]

Os algoritmos de ML têm apresentado aplicações em diferentes áreas, como na previsão dos preços do setor imobiliário, nos filtros de spam, nas redes sociais, no sistema bancário, na medicina, na resposta imunológica de novos fármacos, na educação, no reconhecimento de voz e de imagens. [1]

Na educação, foi recentemente observado que os algoritmos de ML estão sendo utilizados na previsão de notas, na recomendação de cursos de graduação, sistemas de tutoria inteligente, e na melhora do currículo. [3] Pellegrino et al. em 2001 [4] destacaram a importância de melhorar as avaliações fazendo uso das tecnologias emergentes para realmente saber o que os estudantes estão aprendendo.

Uma revisão da literatura da aplicação de ML nas avaliações mostra que os algoritmos são principalmente utilizados para prever as notas dos alunos, explorar as respostas textuais, avaliar os estudantes e apresentar os resultados tanto a professores como estudantes de forma imediata e avaliar o desempenho em jogos medindo as habilidades cognitivas [5]. Estas aplicações foram realizadas criando os algoritmos e utilizando algoritmos comerciais. Ainda, foi observado que os algoritmos supervisionados são majoritariamente utilizados porque não precisam de uma enorme quantidade de dados, como seria necessário para trabalhar com algoritmos não supervisionados. De todos os artigos selecionados, 5 foram da física, 1 da química, 16 da biologia, 19 em ciências gerais, 6 em ciências da terra, 2 em ciências da computação e 1 em ciências médicas.

Zhai em 2021 [6] estabeleceu o termo de “avaliações inovadoras”, referindo-se às práticas avaliativas fazendo uso das tecnologias emergentes. Ele ainda classifica a inovação em 3 níveis: **substituir, transformar e redefinir**. No primeiro, se faz uso da tecnologia apenas para substituir alguma função das avaliações convencionais. No segundo, as avaliações são transformadas de forma que se tem a oportunidade de tomar melhores decisões na instrução dos alunos e tomar ações mais efetivas. Já o terceiro nível – redefinir – indica uma mudança radical na prática avaliativa, onde pode-se abordar desafios fundamentais no aprendizado individual. Este último nível é difícil de ser alcançado porque é necessária uma integração entre tecnologia, teorias educacionais e práticas avaliativas. Zhai destaca que

o ML apresenta um forte potencial para revolucionar a ciência das avaliações porque o princípio do ML é justamente o aprendizado.

Em 2019 no CEDERJ, foi realizada a implementação de uma avaliação online na disciplina de Introdução às Ciências Físicas 1 (ICF1), substituindo as avaliações conhecidas como Oficinas (OF). Estas oficinas tinham por objetivo principal ajudar na aprendizagem do aluno, e isto não estava sendo cumprido. Na oficina antiga, os estudantes faziam o envio da avaliação e as notas eram conhecidas pelos alunos após duas semanas. Este fato não permitia que os alunos pudessem corrigir seus erros em um tempo adequado, além de existir fortes indícios de cópia entre os alunos.

A nova implementação permite que os alunos possam fazê-la no horário desejado, com um tempo máximo de 3 horas e com correção automática, permitindo a verificação imediata dos erros. Além disso, os estudantes podem refazer as avaliações o número de vezes que desejarem, fazendo com que o erro seja parte fundamental no processo de aprendizagem. É interessante ressaltar que nem todos os alunos obtêm a nota máxima e algumas perguntas que surgem a partir disso são: por que esses alunos não atingem a nota máxima? Como podemos dar o suporte adequado, para não só melhorar a nota, mas melhorar os conhecimentos dos alunos? Como identificar a priori quais alunos precisam de uma atenção detalhada antes do semestre acabar minimizando reprovações ou evasões?

Posteriormente, se observou o potencial dos dados que estavam sendo gerados e foi criado um algoritmo em Python que nos permite fazer um seguimento das notas dos alunos e a evolução das mesmas. [7] Esta ferramenta está sendo utilizada na tentativa de ajudar os alunos que não estão conseguindo melhorar as notas.

Inicialmente, esta avaliação tinha sido implementada apenas para dois tópicos dos seis que compõem a disciplina, mas devido à grande aceitação dos alunos (e aos efeitos de isolamento da pandemia), em 2020 foi possível implementar estas avaliações online para toda a ementa da disciplina.

A disciplina de ICF1 conta com 1500 alunos aproximadamente, de 6 cursos diferentes. Com a implementação das novas avaliações online, um grande número de dados está sendo gerado a cada semestre. Assim, este projeto visa implementar algoritmos de ML, supervisionados e não supervisionados, nos dados gerados de 2020-2 a 2021-2. Nosso principal objetivo é dar suporte aos alunos que mais precisam e para isto precisamos ter ferramentas que nos ajudem na predição daqueles que mostrem indícios de uma possível reprovação ou evasão. Para isto, precisamos conhecer os parâmetros que nos ajudem a identificar a estes alunos, que pode não se limitar apenas as notas obtidas nas avaliações. Também, realizaremos uma busca de novas tecnologias utilizando algoritmos de ML que possam ser utilizadas nas nossas avaliações em ICF1.

Cronograma de atividades

A tabela 1 apresenta a lista de atividades a serem desenvolvidas neste projeto. As atividades foram numeradas para facilitar a identificação na tabela 2 que contém o cronograma previsto para o desenvolvimento do trabalho proposto.

Tabela 1: descrição das atividades.

N°	Descrição
1	Limpeza e preparação dos dados para aplicação dos algoritmos
2	Implementação de algoritmos supervisionados
3	Implementação de algoritmos não supervisionados
4	Pesquisa de novos projetos de aprendizado de máquina a ser implementados em ICF1
5	Discussão dos resultados e preparação de apresentações e publicações

Tabela 2: Cronograma de atividades.

Atividade	Mês											
	1	2	3	4	5	6	7	8	9	10	11	12
1	X	X	X									
2			X	X	X							
3						X	X	X				
4								X	X	X		
5										X	X	X

Bibliografia

[1] ALPAYDIN, Ethem. **Introduction to machine learning**. MIT press, 2020.

[2] JANIESCH, Christian; ZSCHECH, Patrick; HEINRICH, Kai. Machine learning and deep learning. **Electronic Markets**, v. 31, n. 3, p. 685-695, 2021.

[3] WEBB, Mary E.; FLUCK, Andrew; MAGENHEIM, Johannes; MALYN-SMITH, Joyce; WATERS, Juliet; DESCHÊNES, Michelle; ZAGAMI, Jason. Machine learning for human learners: opportunities, issues, tensions and threats. **Educational Technology Research and Development**, v. 69, n. 4, p. 2109-2130, 2021.

[4] PELLEGRINO, J. W.; CHUDOWSKY, N.; GLASER, R. (2001). Knowing what students know: the science and design of educational assessment. ERIC

[5] ZHAI, Xiaoming; YIN Yue; PELLEGRINO, James W.; HAUDEK, Kevin C.; SHI, Lehong. Applying machine learning in science assessment: a systematic review. **Studies in Science Education**, v. 56, n. 1, p. 111-151, 2020.

[6] ZHAI, Xiaoming. Practices and theories: How can machine learning assist in innovative assessment practices in science education. **Journal of Science Education and Technology**, v. 30, n. 2, p. 139-149, 2021.

[7] SARMIENTO, Charlie V.; PENELLO, Germano M.; SIGAUD Lucas Mauricio. Análise Quantitativa do Desempenho dos alunos em Avaliações na Disciplina de ICF1 do CEDERJ. **Revista Educação à Distância**, Batatais, v. 10, p. 71-87, jan./jun. 2020