

COD3LYTICS E COD3COLLECTIVE: FERRAMENTA E BASE DE DADOS PARA APLICAÇÃO DE LEARNING ANALYTICS EM PROGRAMAÇÃO COMPETITIVA

ANDRESSA THEOTÔNIO ALVES DE FIGUEIRÊDO (IFPB, Campus João Pessoa), CALEBE OLIVEIRA DE FIGUEIRÊDO (IFPB, Campus João Pessoa), THIAGO GOUVEIA (IFPB, Campus João Pessoa)

Área de conhecimento:(Tabela CNPq): 1.03.03.04-9 Sistemas de Informação.

Palavras-Chave: Learning Analytics; Programação competitiva; Análise de desempenho; Educação computacional; Visualização de dados.

1 Introdução

A participação de estudantes brasileiros em olimpíadas e maratonas de programação vem crescendo de maneira significativa nos últimos anos (FERREIRA, 2024). Tais competições exigem preparo intensivo pois demandam a resolução de problemas complexos sob restrição de tempo e cenário altamente competitivo, estimulando habilidades de raciocínio lógico, abstração de algoritmos e tomada de decisão. Dessa forma, os competidores são expostos à formação de habilidades valorizadas em ambientes profissionais, fazendo da programação competitiva um diferencial durante a formação. Contudo, os competidores precisam conciliar a rotina de treinamentos com os cursos regulares, nos quais a carga curricular nem sempre contempla o aprofundamento necessário em programação (LAAKSONEN, 2022).

Plataformas como AtCoder¹ e Codeforces² são amplamente utilizadas nos treinamentos, oferecendo uma grande base de problemas e julgamento automático de submissões. Porém, o sistema de *feedbacks* dessas plataformas é bastante resumido e restrito a informações básicas, como acertos, erros e *rankings*, sem oferecer *insights* detalhados para o competidor.

Nesse contexto, a aplicação de *Learning Analytics* (LA) surge como alternativa promissora, sendo definida como a coleta, análise e interpretação de dados educacionais com o objetivo de gerar informações acionáveis para melhorar o ensino e a aprendizagem (SoLAR Learning Analytics Task Force, 2025). Esta pesquisa propõe o uso de LA associado ao conceito de Zona de Desenvolvimento Proximal (ZDP), que representa o intervalo entre o que um aprendiz realiza sozinho e o que pode atingir com apoio qualificado. Tecnologias educacionais podem atuar como mediadoras nesse processo, favorecendo intervenções pedagógicas mais contextualizadas (RIGOPOULI; KOTSIFAKOS; PSAROMILIGKOS, 2025). No caso da programação competitiva, essa abordagem permite recomendar problemas com dificuldade levemente superior ao nível atual do competidor, promovendo avanço contínuo.

Em ambientes digitais, como cursos *online* e sistemas educacionais adaptativos, o LA vem sendo utilizado com sucesso para analisar progresso, prever evasão, oferecer *feedback* formativo, sugerir conteúdos personalizados e permitir um apoio à autorregulação da aprendizagem (VESIN et al., 2022). Já aplicada em ambientes educacionais diversos, o uso de *Learning Analytics* ainda é uma estratégia pouco explorada no âmbito de programação competitiva (CHAPARRO; SALINAS; VERGARA, 2021).

Este trabalho apresenta a ferramenta COD3Lytics, que aplica técnicas de LA à análise de desempenho de usuários do Codeforces², gerando métricas descritivas e visualizações interativas. A iniciativa também é sustentada pela construção do dataset COD3Collective, que consolida dados de múltiplas plataformas com foco na padronização e uso futuro em sistemas de recomendação adaptativa.

1.1 Justificativa

A justificativa para o desenvolvimento do COD3lytics surge no ponto em comum entre uma ferramenta capaz de prover análises com base em dados reais e a recomendação de questões aderentes ao nível atual do competidor. A relevância da proposta surge na carência de estudos e ferramentas voltadas para o público de programação competitiva (LUIGI et al., 2018), além da relevância significativa das competições de programação na educação básica e superior (FERREIRA, 2024; LAAKSONEN, 2022).

Entre as plataformas mais utilizadas por competidores, destaca-se o Codeforces, que fornece uma vasta base de problemas e julgamento automático. No entanto, sua interface apresenta limitações relevantes: não há visualizações personalizadas, nem *feedbacks* analíticos sobre desempenho, o que limita a autonomia e o direcionamento estratégico dos estudos dos

¹<https://atcoder.jp/>

²<https://codeforces.com/>

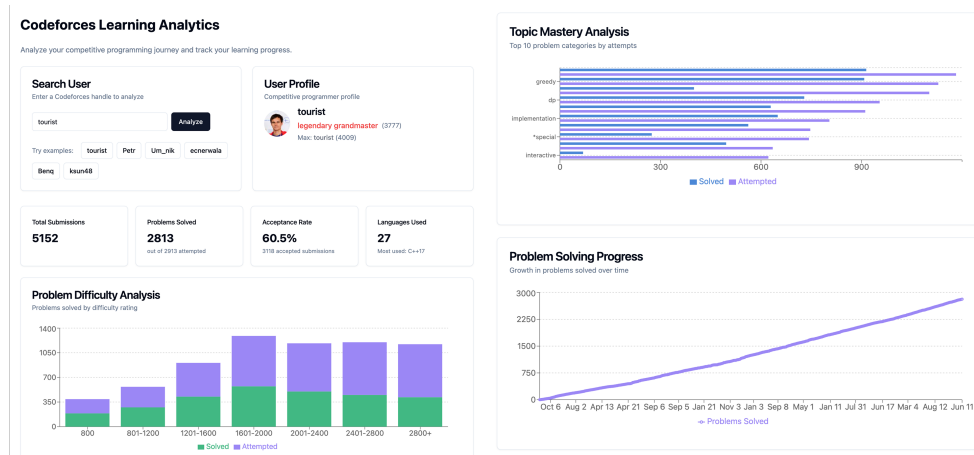


Figura 1: Protótipo inicial do C0D3Lytics.

Da esquerda para a direita: Análise de Dificuldade dos Problemas (*Problem Difficulty Analysis*), Perfil do Usuário (*User Profile*), Domínio por Tópico (*Topic Mastery Analysis*), Progresso de Resolução de Problemas (*Problem Solving Progress*).

usuários. Essa lacuna motivou o desenvolvimento do C0D3lytics para oferecer insights gráficos, agrupamentos por tópicos e métricas temporais, além de possibilitar futuras recomendações inteligentes com base em histórico de submissões.

Dessa forma, o trabalho possui não apenas justificativas de inovação técnica, mas também pelo potencial impacto no campo de educação computacional, promovendo ambientes educativos mais orientados por dados.

O artigo está organizado em três seções principais: a Seção 2 detalha a metodologia adotada, incluindo a construção da ferramenta e do dataset; a Seção 3 apresenta os resultados obtidos até o momento e as funcionalidades implementadas; e a Seção 4 discute as considerações finais e as perspectivas para trabalhos futuros.

2 Materiais e Métodos

A metodologia de desenvolvimento da pesquisa está estruturada em cinco fases principais: (i) coleta de dados, (ii) definição de métricas, (iii) construção dos módulos de análise e visualização, (iv) recomendação de problemas e (v) integração final da ferramenta. O projeto está estruturado de forma a serem trabalhadas duas frentes principais: o desenvolvimento da ferramenta C0D3lytics e a base de dados C0D3Collective, como detalhado nas seções 2.1 e 2.2.

2.1 C0D3lytics: ferramenta de visualização de dados

A ferramenta coleta e gera gráficos a partir de dados de competidores da plataforma Codeforces. A partir do *username* do competidor, o C0D3lytics extrai informações sobre as submissões de cada usuário utilizando a API pública do Codeforces. São extraídos dados como: histórico de submissões, tempo de submissões, acertos, erros e nível de dificuldade das questões submetidas. Com base nesses dados, são apresentados gráficos que mostram métricas descritivas como total de problemas resolvidos, taxa de acertos, distribuição de tópicos (assuntos abordados nas questões), progressão temporal e distribuição de *rating*, como pode ser analisado na figura 1.

2.2 C0D3Collective: dataset de problemas

O dataset está sendo construído para integrar problemas de diversas plataformas, como Codeforces², Kattis³, AtCoder e G4M3⁴. Atualmente, a coleta de dados está centrada nas primeiras três plataformas, sendo realizada através da API pública para o Codeforces, via API da comunidade (projeto Kenkoooo⁵) para o AtCoder e via *scraping* na base AutoKattis para o Kattis.

³<https://open.kattis.com/>

⁴<http://g4m3.c073.com/>

⁵<https://github.com/kenkoooo/AtCoderProblems/blob/master/doc/api.md>

Diferentemente da coleta para a C0D3lytics, o foco da coleta de dados para o dataset está nos problemas presentes nas plataformas. Os dados coletados são o título do problema, nível de dificuldade (*rating*), tópicos associados, URL e número de resoluções. Os dados estão unificados no formato .csv com identificadores únicos para cada problema. Atualmente, o dataset conta com 21.181 problemas integrados.

A fase atual está concentrada na normalização das dificuldades (*rating*) de cada problema para o padrão do Codeforces, já que cada plataforma possui escalas diferentes para classificar a dificuldade das questões. Além disso, será testado o uso de um modelo de linguagem natural (LLM) para rotular os problemas por tópico, possibilitando utilização no algoritmo de recomendação.

3 Resultados e Discussão

O protótipo do C0D3lytics está funcional, apresentando visualizações e métricas para facilitar o entendimento do desempenho individual do competidor em um cenário geral de submissões. Os dados estão sendo apresentados visualmente e agrupados em tópicos, dificuldade e evolução do competidor, permitindo análises básicas sobre padrões e gargalos nos treinamentos ao passar do tempo.

O dataset C0D3Collective avança na normalização de *rating* e rotulagem de tópicos, buscando prover a aplicação de algoritmos de recomendação baseados na ZDP e no *rating* (VESIN et al., 2022) adaptado ao contexto de programação competitiva e educação computacional. Além disso, podem ser feitas análises exploratórias e estatísticas para comparação de problemas entre plataformas. Dentro da fase inicial de desenvolvimento, os resultados demonstram viabilidade técnica e potencial de inovação no ecossistema de educação em programação competitiva.

4 Considerações Finais

O projeto propõe o uso de Learning Analytics na programação competitiva para acompanhar treinamentos e oferecer recomendações personalizadas com base no desempenho dos competidores. A ferramenta C0D3lytics já realiza análises sobre usuários do Codeforces, com visualizações e métricas que ajudam a entender sua evolução.

Como próximos passos, prevê-se o estudo de iniciativas semelhantes, a ampliação das métricas, melhorias na interface e funcionalidades da plataforma, além do desenvolvimento de um modelo de recomendação de problemas que considere o nível do competidor com base na Zona de Desenvolvimento Proximal (ZDP) e no C0D3Collective. A proposta também inclui validar a ferramenta em contextos educacionais reais, investigando seu impacto no desempenho, engajamento e autonomia, buscando integrar tecnologia, educação e formação em computação com foco no uso de dados na programação competitiva.

Referências

- CHAPARRO, E.; SALINAS, P.; VERGARA, M. Learning analytics in computer programming courses: A systematic literature review. *Informatics in Education*, v. 20, n. 2, p. 275–301, 2021.
- FERREIRA, C. Competições científicas. In: *Relatório Anual da SBC 2023–2024*. Porto Alegre: Sociedade Brasileira de Computação, 2024. p. 27–29. Acesso em: 13 jun. 2025. Disponível em: <<https://books-sol.sbc.org.br/index.php/sbc/catalog/book/144>>.
- LAAKSONEN, A. What is the competitive programming curriculum? *Olympiads in Informatics*, IOI, Vilnius University, v. 16, p. 35–42, 2022. Disponível em: <https://ioinformatics.org/journal/v16_2022_35_42.pdf>.
- LUIGI, W. D. et al. Learning analytics in competitive programming training systems. In: IEEE. *22nd International Conference on Information Visualisation (IV)*. [S.l.], 2018. p. 321–328.
- RIGOPOULI, K.; KOTSIFAKOS, D.; PSAROMILIGKOS, Y. Vygotsky's creativity options and ideas in 21st-century technology-enhanced learning design. *Education Sciences*, MDPI, v. 15, n. 2, p. 257, 2025. Disponível em: <<https://www.mdpi.com/2227-7102/15/2/257>>.
- SoLAR Learning Analytics Task Force. *Reimagining Learning Analytics*. [S.l.], 2025. Accessed: 2025-06-14. Disponível em: <<https://www.solaresearch.org/wp-content/uploads/2025/06/Reimagining-Learning-Analytics-V3-002.pdf>>.
- VESIN, B. et al. Adaptive assessment and content recommendation in online programming courses: On the use of elo-rating. *ACM Transactions on Computing Education (TOCE)*, ACM, v. 22, n. 3, p. 1–27, 2022. Disponível em: <<https://doi.org/10.1145/3511886>>.