

Blockly-DS: Programação em Blocos para o ensino de Ciência de Dados

Luiz Barboza¹, Rafael Ferreira Mello¹, Erico Teixeira¹

¹CESAR.School Recife-PE Brasil

{lcbj, rfm, est}@cesar.school

Resumo. O interesse pela ciência de dados vêm crescendo nos últimos anos em todas as áreas, sejam STEM ou não-STEM. Alunos não-STEM muitas vezes apresentam dificuldades em programação, mas tem facilidade na absorção ao se utilizar ferramentas visuais. Assim, para este grupo específico utilizar ferramentas vêm se demonstrando essencial para o ensino de ciência de dados. Existem diversas ferramentas que estão sendo propostas. Porém todas apresentam algum tipo de limitação. Desta forma, o objetivo deste trabalho é apresentar uma nova ferramenta, chamada de Blockly-DS, capaz de auxiliar no ensino de ciência de dados para o público não-STEM. A ferramenta Blockly-DS é um protótipo que está em fase de testes em duas instituições brasileiras, no IBMEC, voltado a alunos de graduação, e na Lumturo, cursinho pré-vestibular.

Abstract. Interest in data science has been growing in recent years in all areas, whether STEM or non-STEM. Non-STEM students often have difficulties in programming, but are easily absorbed when using visual tools. Thus, for this specific group, the use of tools has been shown to be essential for the teaching of data science. There are several tools that are being proposed. However, they all have some kind of limitation. In this way, the objective of this work is to present a new tool, called Blockly-DS, capable of assisting in the teaching of data science to the non-STEM audience. The Blockly-DS tool is a prototype that is being tested at two Brazilian institutions, at IBMEC, aimed at undergraduate students, and at Lumturo, a pre-university course.

1. Problema pedagógico

À medida que a ciência de dados se torna difundida em todos os campos, os currículos introdutórios necessitam de novas ferramentas para motivar alunos novatos no que diz respeito à ciência de dados, inteligência artificial (IA) e aprendizado de máquina (do inglês *Machine Learning* - ML). Assim, a formação técnica passa a ser importante para todas as áreas, seja STEM (estatística, tecnologia, engenharia e matemática) ou não-STEM. Entretanto, os alunos não-STEM inicialmente apresentam dificuldades na realização de tarefas voltadas à ciência de dados, principalmente quando existe a necessidade da utilização de programação. Desta forma, ferramentas de apoio ao ensino podem diminuir as dificuldades desses alunos. O objetivo da implementação de ferramentas de apoio ao ensino de ciência de dados para não-STEM é melhorar os currículos, enriquecendo-os com conteúdo autêntico do dia a dia, em estruturas que podem guiar os alunos na resolução dos problemas de suas áreas específicas a partir destas ferramentas automatizadas. Um exemplo das

ferramentas que podem auxiliar os alunos não-STEM são as linguagens baseadas em blocos, como o Google Blockly [Pasternak et al. 2017]. Nesta ferramenta, alunos com pouco ou nenhum conhecimento conseguem iniciar a escrita de códigos de forma conveniente.

2. Contexto educacional

O Blockly-DS é uma extensão para a ferramenta de linguagem de programação visual Blockly [Pasternak et al. 2017], originalmente desenvolvida pelo Google. Esta extensão foi desenvolvida como parte do trabalho de doutorado desenvolvido na CESAR School. A ferramenta tem sido então aplicada primariamente no ensino de ciência de dados para alunos não-STEM, particularmente nos cursos de graduação em economia e administração da escola de negócios de São Paulo, IBMEC e na Lumturo, cursinho popular online voltado para o ingresso na universidade. Além disso, ferramentas visuais podem facilitar esse processo. A ferramenta base estendida aqui, o Blockly, que foca primariamente no ensino de conceitos básicos de programação como variáveis, funções, ifs e loops, permitindo a execução de programas visuais, bem como a geração de código em linguagens como JavaScript, Python, PHP e outras. O Blockly-DS tem como objetivo trazer funcionalidades de análises preditivas trabalhando algoritmos clássicos de IA supervisionados e não-supervisionados, para classificação, regressão e associação, entre eles: Redes Neurais, Regressão Linear, Regressão Logística, KNN, Naive Bayes, Árvore de Decisão e K-Means. Tal qual, métricas como Acurácia, Precisão, *Recall*, Matriz de Confusão e Score.

3. Abordagem pedagógica

Estudos demonstram [Ogegbo and Ramnarain 2022] que o uso ponderado das ferramentas computacionais podem aprofundar o aprendizado de ciência de dados de uma perspectiva pedagógica, pois as ferramentas fornecem um contexto relevante suficientemente poderoso capaz de criar soluções completas para as questões as quais os alunos são submetidos [Gresse von Wangenheim et al. 2021]. As ferramentas juntamente com as abordagens instrucionais criativas, como programas baseados em modelos ou PBL, podem promover a compreensão do conteúdo e o desenvolvimento de habilidades relacionadas à ciência de dados, CT, IA e ML.

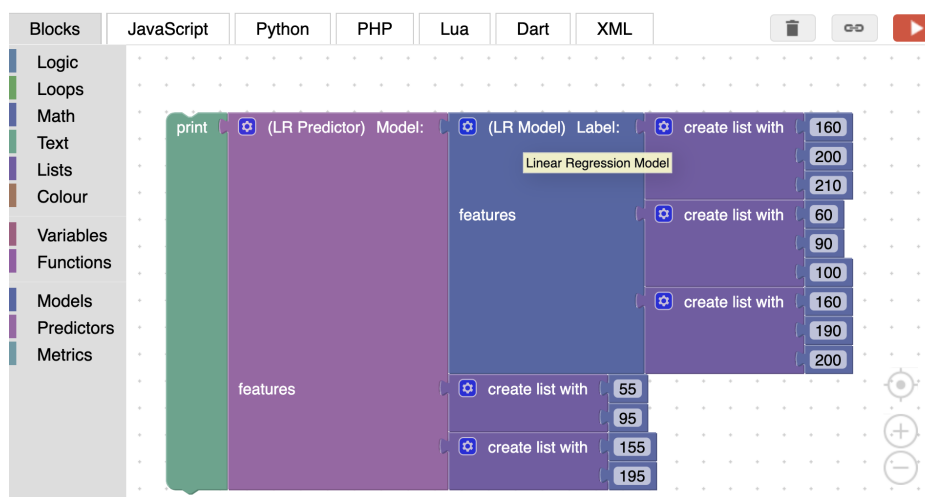
4. Inovação apresentada

Atualmente existem variações da ferramenta do Google Blockly, assim ferramentas similares estão sendo criadas. Porém, muitas destas ferramentas similares apresentam algum tipo de limitação. Como exemplo têm-se o Blockly [Xu et al. 2020]. O Blockly é uma plataforma de código aberto que possui um contexto de ciência de dados que permite aos alunos se conectarem com o conteúdo educacional em cenários do mundo real por meio de problemas relevantes para a sociedade. Esta ferramenta foca na geração de códigos Python a partir de blocos, trazendo elementos específicos do Python, diferentemente da ferramenta original do Google. A introdução do Python no Blockly ocorre, pois, o Python é uma linguagem popular, amigável para iniciantes. Todavia, o Blockly não é a melhor ferramenta para se utilizar como apoio ao ensino de ciência de dados para não-STEM, pois a única funcionalidade que se aproxima de uma capacidade analítica, são bases de dados pré-disponibilizadas que podem ser carregadas. Outro exemplo é a integração do Blockly ao Jupyter, que cria a extensão JupyterLab [Bisong 2019]. O sucesso do Projeto

Jupyter integrou diversas plataformas, e hoje possuem algumas variações como o JupyterLab. Quando o usuário seleciona a extensão JupyterLab, ocorre o sincronismo entre códigos Python e a visualização gráfica em blocos. O JupyterLab tem se tornado muito utilizado, pois apresenta um ambiente de codificação que não requer software externo, possibilitando a utilização de códigos embutidos e é particularmente adequado para o ensino de ciência de dados, IA e ML.

5. Apresentação do software

O Blockly-DS pode ser utilizado e acessado diretamente via um navegador de internet, através do endereço: https://quant-ibmec.group/blockly_ds. Neste ambiente tem-se todas funcionalidades usuais de conceitos básicos de programação como variáveis, funções, ifs e loop disponíveis no Blockly do Google. A extensão ciência de dados (DS) proposta disponibiliza três novos menus, *Models* (Modelos), *Predictors* (Preditores) e *Metrics* (Métricas). Nos modelos é possível a execução de qualquer um dos modelos de IA para ser ajustado (*fit*) com dados de treinamento, tipicamente representados na forma de listas numéricas, conforme visto na figura 1.



Uma vez treinado o modelo, é possível medir o nível de acerto com o *Score* para um classificador e *R2* para um regressor. Estes blocos de códigos conectados podem ser executados em tempo real, através de um botão de execução (*Play*) no canto superior direito. Estando em um nível de acerto aceitável, é possível conectar o preditor correspondente, atrelando novos dados (denominados dados de teste). Este preditor pode então inferir novos dados com base nestes dados de teste. Além da própria inferência, é possível realizar várias mensurações tanto para regressores quanto para classificadores conforme detalhada na seção anterior. Além da execução de blocos lógicos interconectados em tempo real, é possível gerar códigos de programação primariamente para as linguagens JavaScript e Python. Os códigos Python podem ser executados no ambiente de linguagem de preferência do usuário. Neste trabalho tipicamente recomenda-se o uso do Google Colab, por não necessitar de qualquer instalação ou configuração, permitindo o uso de ambiente em nuvem. Os detalhes desta utilização podem ser vistos no material complementar disponibilizado, em Instruções de Uso.

6. Impacto no processo educacional

Blockly-DS foi desenvolvido em JavaScript, estendendo o Blockly com menus de Modelos, Preditores e Métricas. Além da execução de códigos em tempo real, a ferramenta permite a geração automática de código fonte nas linguagens JavaScript e Python. Como funcionamento, os modelos recebem dados de treino, os quais são utilizados para ajuste do modelo (*fit*), podendo assim ser utilizado como base de previsão de novos dados. A primeira versão do Blockly-DS tem como objetivo trazer funcionalidades de análises preditivas trabalhando algoritmos clássicos da IA (supervisionados e não-supervisionados) e algoritmos de ML, para classificação, regressão e associação, sendo eles: regressão linear, regressão logística, KNN, naive bayes, árvore de decisão, redes neurais e K-Means. Os elementos de modelos de IA então podem ser utilizados em conjunção com seu respectivo preditor, para realizar previsões com base em um novo conjunto de dados. Desta maneira tal qual os modelos, os preditos podem ser categorizados em regressores supervisionados, classificadores supervisionados e algoritmos de associação não-supervisionados. Tanto para os modelos treinados como dados de treinamento, quanto para preditores conectados a modelos que podem realizar previsões com base em novos dados, por exemplo dados de teste, estes podem ter seu desempenho mensurado. Estas mensurações se aplicam usualmente a algoritmos supervisionados, tanto regressores quanto classificadores.

7. Considerações finais

A ferramenta por esta em fase de conclusão de desenvolvimento da sua primeira versão, ainda não foram coletados dados da eficiência de uso desta. Os primeiros de grupos de teste e controle de utilização da ferramenta esta agendado para Julho de 2022 na escola Lumturo. A Lumturo é um cursinho popular online voltado para todos e todas que desejam ingressar em um universidade. Como conteúdo complementar, a Lumturo oferece curso de verão de diversos temas, entre eles de tecnologia, como introdução a programação e ciência de dados.

Instruções: quant-ibmec.group/blockly_ds/Instrucoes.pdf

Vídeo de demonstração: youtube.com/watch?v=H60EfhBduCo

Referências

- Bisong, E. (2019). Jupyterlab notebooks. In *Building Machine Learning and Deep Learning Models on Google Cloud Platform*, pages 49–57. Springer.
- Gresse von Wangenheim, C., Hauck, J. C., Pacheco, F. S., and Bertoneceli Bueno, M. F. (2021). Visual tools for teaching machine learning in k-12: A ten-year systematic mapping. *Education and Information Technologies*, 26(5):5733–5778.
- Ogegbo, A. A. and Ramnarain, U. (2022). A systematic review of computational thinking in science classrooms. *Studies in Science Education*, 58(2):203–230.
- Pasternak, E., Fenichel, R., and Marshall, A. N. (2017). Tips for creating a block language with blockly. In *2017 IEEE blocks and beyond workshop (B&B)*, pages 21–24. IEEE.
- Xu, C., Feng, Z., Qi, P., and Sun, Y. (2020). An automatic analysis tool based on computational thinking for blockpy programs. In *2020 International Wireless Communications and Mobile Computing (IWCMC)*, pages 2045–2050. IEEE.