

Aplicação da Ferramenta Google Colaboratory no Ensino de Ciências de Dados

Martony Demes da Silva¹

¹Departamento de Computação – Universidade Federal do Piauí (UFPI)
Teresina – PI – Brazil

Abstract. *This paper presents an evaluation of the Google Collaboratory environment (or colab) to aid in the teaching of a student of Data Science training. This tool facilitates student interaction in the practice of programming activities, application of Machine Learning techniques and Artificial Intelligence. To validate this proposal, an analysis of perceived utility and perceived ease of use of the tool was used. Results show that in 10 of the 12 questions, they presented an index above 90%*

Resumo. *Esse trabalho apresenta uma avaliação do ambiente Google Colaboratory (ou colab) para o auxílio no ensino de estudante de um treinamento de Ciências de Dados. Essa ferramenta facilita a interação do aluno na prática de atividades de programação, aplicação de técnicas de Aprendizagem de Máquina e Inteligência Artificial. Para validação desta proposta, utilizou-se de uma análise de utilidade percebida e facilidade de uso percebida da ferramenta. Resultados mostram que em 10, das 12 questões, apresentaram índice acima de 90%.*

1. Introdução

O ambiente *Google Colaboratory* ou *Colab* é um serviço de nuvem gratuito hospedado pelo *Google* para incentivar a pesquisa de Aprendizado de Máquina e Inteligência Artificial [Carneiro et al. 2018]. Por meio dos recursos disponíveis, *Colab* facilita o processo de programação e prática com Ciências de Dados. Dada essa praticidade, ela também é utilizada em outras disciplinas como na matemática [Access 2020] e medicina [Balaraman 2020].

Tomando algumas pesquisas relacionadas [Neto and Schuvartz 2007], [Amaral et al. 2017] e [Blatt et al. 2017], verificou-se que o ensino e prática da disciplina de programação ainda é um desafio para professores de computação. Os estudantes iniciantes precisam adquirir habilidades como abstração e raciocínio lógico. Os professores, por outro lado, se esforçam para contornar isso. Porém ainda há um alto índice de reprovação em programação [Amaral et al. 2017].

Outro desafio existente, é a construção e manutenção do ambiente computacional para práticas de programação. Na maioria dos casos, para iniciar um curso ou disciplina de programação deve-se preparar e manter um complexo conjunto de ferramentas e pacotes de bibliotecas.

Dado este cenário, esta pesquisa engajou-se sobre o ensino de ciências de dados (ou *Data Science*). A aprendizagem de Ciências de dados requer estudar as informações, o processo de captura, transformação, geração e, posteriormente, análise de dados.

Essa aprendizagem é interdisciplinar e envolve: estatística, computação, conhecimento do negócio, matemática entre outras [Grus 2019]. Especificamente, para se ter boas práticas com ciências de dados, é importante que o estudante tenha conhecimentos sobre programação.

Além disso, o uso de ferramentas colaborativas possibilita novas perspectivas ao processo de ensino-aprendizagem. Desta forma, ressalta-se que é possível prover novas maneiras de realizar as atividades educacionais, incluindo planejamento integrado e funcionalidades específicas, em que os envolvidos possam executar as tarefas na internet de modo colaborativo [Abegg et al. 2010], [Mattos et al. 2017] e [Delbem et al. 2019].

Nesse contexto, este trabalho objetiva avaliar a ferramenta de desenvolvimento online e colaborativa *google colab* no âmbito do ensino de Ciências de Dados. A utilização deste ambiente visa facilitar a interação do aluno na prática da disciplina por meio dos diversos recursos disponíveis na plataforma *colab*. A ferramenta também possibilita que vários alunos interajam colaborativamente. Para validar esta pesquisa, foi feita uma avaliação com estudantes do curso de *Data Science*, na prática.

Vale ressaltar que, de fato, há muitas metodologias de ensino de programação que usam diversos tipos de ferramentas para apoiar os vários níveis educacionais de criança, jovens e adultos. Nesse sentido, o diferencial desta pesquisa é uso específico da ferramenta *Google Colab* e no contexto de ensino de ciências de dados.

Este trabalho está organizado da seguinte forma: na Seção 2 abordou-se sobre trabalhos relacionados ao assunto em comento; A Seção 3 apresenta a ferramenta do estudo; Na Seção 4 é feita a metodologia da pesquisa; A Seção 5 ilustra os resultados da pesquisa e a Seção 6 finaliza com as considerações finais.

2. Referencial Teórico

Como citado na Seção 1, o estudo de Ciências de Dados envolve aplicação multidisciplinar [Grus 2019] no processo de analisar dados para um dado contexto. Essa análise envolve capturar, transformar, gerar e avaliar os dados. Por exemplo, a partir de um conjunto de dados de milhares de pacientes é possível mapear predisposição para determinada doença.

Entre as habilidades necessárias para trabalhar com Ciências de Dados está a prática em programação. Diversos trabalhos foram encontrados sobre os desafios com o ensino-aprendizagem de programação. Diante disso, uma proposta para transpor as dificuldades no ensino de programação é o uso de ferramentas de inteligência artificial [Neto and Schuvartz 2007].

O mesmo problema é abordado pelos pesquisadores em [Farias et al. 2015]. No bojo desse trabalho, cita-se a falta de conhecimento nas funcionalidades das ferramentas de programação e na pouca intuitividade de suas interfaces. Como evolução, eles propõem ferramenta que facilita a comunicação e interação entre estudantes iniciantes e professor.

No encadeamento teórico relacionado, outras pesquisas propuseram e validaram o uso de ferramenta de apoio ao processo de ensino e aprendizagem de lógica e programação para iniciantes como em [Amaral et al. 2017].

De modo suplementar, destaca-se o trabalho de mapeamento sistemático sobre

metodologias e ferramentas para apoio ao ensino de programação, de [Blatt et al. 2017]. O principal enfoque dessa pesquisa é sobre a preferência dos professores em utilizar *Scratch* como ferramenta de apoio. O *Scratch* possibilita que estudantes (inclusive crianças) elaborem blocos visuais por meio de animações, histórias interativas ou jogos.

No âmbito do tema pesquisa, não encontrou-se trabalhos com uso de ferramentas aplicadas ao ensino de Ciências de Dados. Além disso, os trabalhos sobre uso da ferramenta *Google colab* encontrados restringiram-se a: comparar com uso de um hardware real [Carneiro et al. 2018], uso da ferramenta para comparar algoritmos de classificação [Balaraman 2020] e aplicação em estudo matemático [Alves and Machado Vieira 2019] e [Access 2020].

3. Google Colaboratory - Colab

O *Google Colab* ou “*Colaboratório*” é um serviço de nuvem gratuito hospedado pelo Google para incentivar a pesquisa de Aprendizado de Máquina e Inteligência Artificial [Vishakha Lall 2018]. O ambiente digital é uma máquina virtual, ou notebook, em que o estudante pode desenvolver práticas de programação em python e aplicar os conceitos de Ciências de Dados.

Os benefícios do colab são:

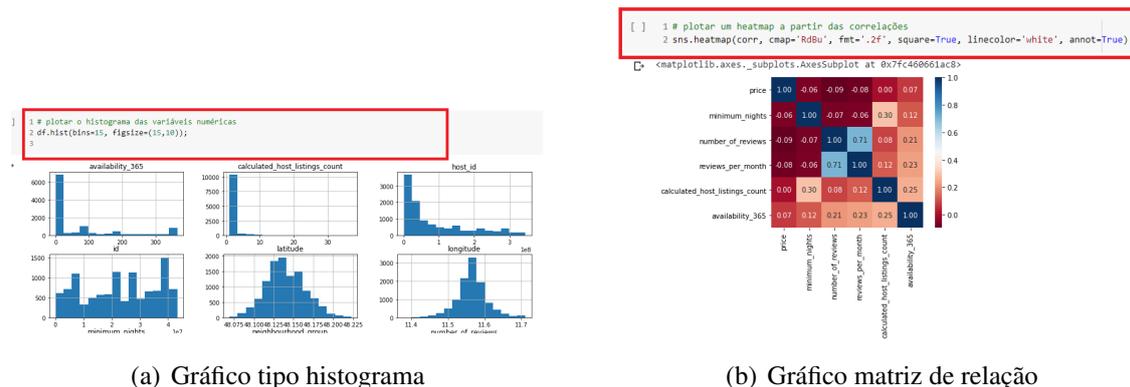
- Suporte para linguagem Python;
- Aceleração de GPU grátis;
- Bibliotecas pré-instaladas: Todas as principais bibliotecas Python, como o TensorFlow, o Scikit-learn, o Matplotlib, entre muitas outras, estão pré-instaladas e prontas para serem importadas [Vishakha Lall 2018];
- Construído com base no Jupyter Notebook;
- Recurso de colaboração (funciona com uma equipe igual ao Google Docs): permite que os desenvolvedores usem e compartilhem o Jupyter notebook entre si sem precisar baixar, instalar ou executar qualquer coisa que não seja um navegador;
- Suporte a comandos bash;
- Os notebooks do Google Colab são armazenados no drive.

Uma característica importante do *colab*, é a possibilidade distribuir separadamente blocos de texto e de código dentro do mesmo arquivo. Isso favorece o suporte a comentários do código, compilação separada, elaboração simultânea de relatório, clareza e organização dos estudos. Essas características são essenciais em Ciências de Dados. A Figura 1, a seguir, apresenta um exemplo de bloco de código e de bloco de texto.



Figura 1. Obtenção e análise de dados

Outro fator relevante, visualmente, a ferramenta favorece que o estudante na elaboração de gráficos e tabelas. A Figura 2 ilustra exemplos de gráficos criados no *colab*. Na parte de cima de cada imagem, destacado em vermelho, apresenta o código inserido e em seguida pode ser compilado dentro deste bloco, sem afetar outras parte do arquivo (ou *notebook*).



(a) Gráfico tipo histograma

(b) Gráfico matriz de relação

Figura 2. Exemplos de gráficos no colab

Como citado no trabalho de [Carneiro et al. 2018], dado um ambiente controlado, o *colab* apresenta uma performance compatível com servidores reais. Além disso, devido sua interface amigável, é possível ser utilizado por qualquer estudante mesmo que não tenha experiência com programação. Na Seção 4 será apresentada uma avaliação realizada no *colab* com estudantes do curso de Ciências de Dados.

4. Metodologia

Como forma de validar esta pesquisa, fez-se uma avaliação por meio de um experimento quantitativo com alunos do curso de *Data Science*. A avaliação foi feita com a aplicação de um questionário com 21 alunos. Fez-se um levantamento inicial sobre esses alunos com as seguintes perguntas: 1) Qual sua idade? 2) Tem formação superior em Tecnologia da Informação e 3) Qual sua experiência com programação. Com base nas respostas dessas questões, estruturou-se algumas análises. Verificou-se, conforme gráfico da Figura

3, que 52% não é da área de Tecnologia da Informação. Apenas 28,6% já é formado na área de TI. Outra informação relevante é que cerca de 43% dos envolvidos não tem nenhuma experiência com programação.

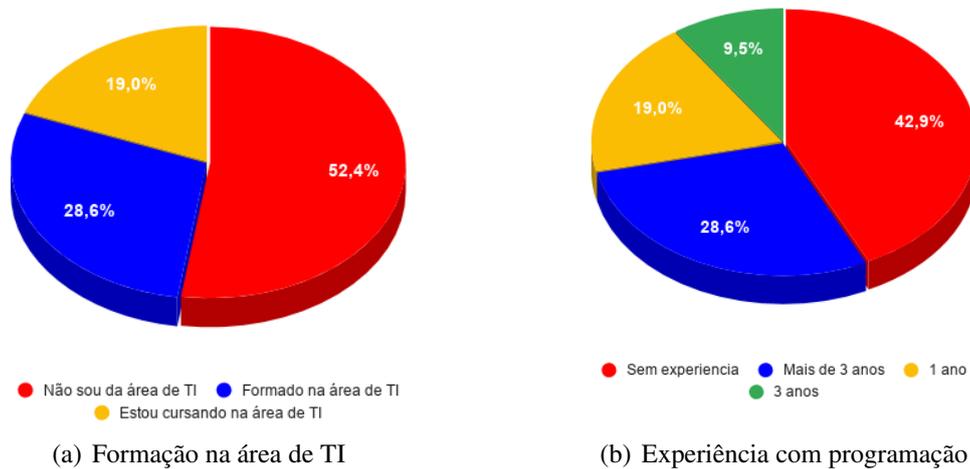


Figura 3. Informações sobre os participantes

Após o levantamento inicial, aplicou um questionário com base na escala Likert [Wainerman 1976] e que seguiu a metodologia *Technology Acceptance Model (TAM)* [Davis et al. 1989]. O TAM objetiva apresentar as intenções de uso de um sistema. Essas intenções são fundamentadas em dois aspectos: **utilidade percebida**, que mede o grau em que um indivíduo acredita que o uso do sistema pode melhorar suas atividades; e **facilidade de uso percebida**, que mede o grau que o usuário acredita que o uso do sistema de informação livre de esforço.

A escala *Likert* verifica o nível de concordância do sujeito com várias afirmações que expressam algo favorável ou desfavorável em relação a um objeto psicológico. Baseado na metodologia utilizada no trabalho de [Neto et al. 2018], os participantes da pesquisa vão escolher entre opções e marcando a resposta conforme sua atitude ou opinião. As opções de respostas disponíveis para cada questão, nesta pesquisa, são: Discordância Total (DT); Discordância Parcial (DP); Discordância Leve (DL); Neutro (N); Concordância Leve (CL); Concordância Parcial (CP); Concordância Total (CT).

As respostas foram analisadas considerando dois pontos [Macnaughton 1996]: quantidade de concordantes (*ConP*) e Discordantes (*DisP*). *ConP* é o cálculo da soma entre a quantidade de respostas totalmente concordantes, parcialmente concordantes, concordância Leve e neutro dividida por 2 (Equação 1). Já o *DisP* é calculado da soma entre a quantidade de respostas totalmente discordantes, parcialmente discordantes, discordância leve e neutro (Equação 2).

$$ConP = CT + CP + CL + \frac{N}{2} \quad (1) \quad DisP = DT + DP + DL + \frac{N}{2} \quad (2)$$

A partir do resultado de *ConP* e *DisP*, calcula-se o grau de concordância de cada

proposição (GCP), pela Equação 3, com base na abordagem de [Wilder 1978].

$$GCP = 100 - \left(\frac{100}{\frac{ConP}{Disp} + 1} \right) \quad (3)$$

Diante dos resultados obtidos de GCP, utiliza-se a Tabela 1 de referência de valores de GCP da qual indica o quanto o participante concorda ou discorda sobre cada questão [Davis et al. 1989].

Tabela 1. Referência de valores do GCP

Valor de GCP	Frase adequada
90 ou mais	Uma concordância muito forte
80 a + 89,99	Uma concordância substancial
70 a + 79,99	Uma concordância moderada
60 a + 69,99	Uma concordância baixa
50 a + 59,99	Uma concordância desprezível
40 a + 49,99	Uma discordância desprezível
30 a + 39,99	Uma discordância baixa
20 a + 29,99	Uma discordância moderada
10 a + 19,99	Uma discordância substancial
9,99 ou menos	Uma discordância muito forte

Diante do exposto, o delineamento exposto nesta pesquisa objetiva avaliar a facilidade de uso da ferramenta. Vale frisar que não foram encontrados na literatura trabalhos semelhantes e assim comparar com os resultados deste trabalho. Porém, a abordagem aplicada pode ser replicável em cenários semelhantes ou em análise em parâmetros diferentes (com mais participantes por exemplo) em pesquisas futuras.

5. Resultados

O resultado do questionário baseado na abordagem de [Neto et al. 2018], leva em consideração dois aspectos: a **Utilidade Percebida**, que constitui 6 proposições e a **Facilidade de Uso Percebida**, compondo também 6 proposições. A Tabela 2 ilustra as questões definidas para esta pesquisa.

Tabela 2. Proposições da Pesquisa

Utilidade percebida	
nº	Questão
1	A utilização do <i>Colab</i> é importante para o meu aprendizado?
2	As funcionalidades do <i>Colab</i> oferecem condições semelhantes a um ambiente de IDE (Ambiente de desenvolvimento Integrado)?
3	O <i>Colab</i> facilita meu entendimento de assuntos passados em aula?
4	Usar o <i>Colab</i> agrega valor ao meu aprendizado em <i>Data Science</i> ?
5	As universidades deveriam utilizar o <i>Colab</i> para treinamento de alunos de curso de computação?
6	Estou motivado a continuar usando o <i>Colab</i> ?
Facilidade de uso percebida	
7	No <i>Colab</i> eu sempre sei onde estou e como chegar onde quero chegar?
8	O <i>Colab</i> tem uma interação compreensível e clara?
9	Os recursos de programação e navegação são fáceis de encontrar?
10	O <i>Colab</i> possui uma boa amigabilidade (fácil de aprender a usar)?
11	Consigo utilizar o <i>Colab</i> sem auxílio de um instrutor?
12	Utilizar o <i>Colab</i> é agradável?

Dada a tabela de questões da pesquisa, realizou-se os cálculos de respostas de 21 participantes da pesquisa. O resultado está apresentado na Tabela 3.

Analisando a Tabela 3, no quadro de **Utilidade Percebida**, observa-se que quatro das seis questões tem *Uma concordância muito forte* (levando como parâmetro a Tabela 1), com índice GCP acima de 90%. E há duas questões com GCP entre 80% a 89,99%, enquadrando em *Uma concordância substancial*. A partir desses resultados, avalia-se que o *Colab* apresenta uma boa utilidade percebida pelos estudantes.

A Facilidade de Uso Percebida verifica o grau no qual usuário acredita que o uso da ferramenta é livre de esforço. O resultado ilustrado na Tabela 3 apresenta em todas as questões um valor GCP acima de 90%. Isso evidencia que o *Colab* é de fácil utilização e sem necessidade de esforço.

Tabela 3. Resultados da Pesquisa

Utilidade percebida											
Questão	DT	DP	DL	N	CL	CP	CT	QTR	DisP	ConP	GCP
1	0	0	0	3	4	4	10	21	1,5	19,5	92,86
2	1	0	2	2	3	8	5	21	4	17	80,95
3	0	0	0	2	4	4	11	21	1	20	95,24
4	0	0	0	3	1	4	13	21	1,5	19,5	92,86
5	0	2	0	1	2	3	13	21	2,5	18,5	88,10
6	0	0	0	1	1	5	14	21	0,5	20,5	97,62
Facilidade de uso percebida											
Questão	DT	DP	DL	N	CL	CP	CT	QTR	DisP	ConP	GCP
7	0	0	0	3	5	7	6	21	1,5	19,5	92,86
8	0	0	0	2	5	4	10	21	1	20	95,24
9	0	0	0	2	3	6	10	21	1	20	95,24
10	0	0	0	1	4	5	11	21	0,5	20,5	97,62
11	0	0	0	2	5	4	10	21	1	20	95,24
12	0	0	0	1	2	4	14	21	0,5	20,5	97,62

Uma análise importante extraída desta pesquisa é que, apesar de 52,4% informar que não é da área de Tecnologia da Informação (TI) e 42,9% informarem que não tem experiência com programação, ilustrados na Tabela 2. Comparando com os resultados acima apresentados, isso reforça que a ferramenta *Colab* é facilitadora no ensino de ciências de dados inclusive para quem não é da área de TI.

Como citado, não foi encontrado pesquisas com uso de ferramentas computacionais no ensino de Ciências de Dados. Dada essa lacuna, este trabalho apresentou resultados relevantes no uso da ferramenta *Colab*. Esses resultados representam uma potencial utilidade dessa ferramenta dentro do contexto estudado, no estudo de Ciências de Dados.

6. Considerações Finais

Esta pesquisa avaliou um ambiente colaborativo *Google Colab* como ferramenta de apoio ao ensino de Ciências de Dados. Os resultados, citados na Seção 5, apresentam que a utilização deste software facilita a prática de programação e de manipulação de dados no contexto de Ciências de Dados.

Nesse contexto, o estudante pode concentrar-se em evoluir nas práticas relacionadas à ciências de dados - estudar as informações, o processo de captura, a transformação, geração e, posteriormente, análise de dados. Assim, esta pesquisa contribui para os avanços na área de estudo e ensino de Ciências de Dados;

Para trabalhos futuros, pode-se fazer comparativo de performance de processamento do *colab* com outras ferramentas relacionadas.

Referências

- [Abegg et al. 2010] Abegg, I., Bastos, F. d. P. d., and Müller, F. M. (2010). Ensino-aprendizagem colaborativo mediado pelo wiki do moodle. *Educar em Revista*, 1(38):205–218.

- [Access 2020] Access, O. (2020). Visualizing the Newtons Fractal from the Recurring Linear Sequence with Google Colab : An Example of Brazil X Portugal Research. 15(3).
- [Alves and Machado Vieira 2019] Alves, F. R. V. and Machado Vieira, R. P. (2019). The Newton Fractal's Leonardo Sequence Study with the Google Colab. *International Electronic Journal of Mathematics Education*, 15(2).
- [Amaral et al. 2017] Amaral, E., Camargo, A., Gomes, M., Richa, C. H., and Becker, L. (2017). ALGO+ Uma ferramenta para o apoio ao ensino de Algoritmos e Programação para alunos iniciantes. *Anais do XXVIII Simpósio Brasileiro de Informática na Educação (SBIE 2017)*, 1(Cbie):1677.
- [Balaraman 2020] Balaraman, S. (2020). Comparison of Classification Models for Breast Cancer Identification using Google Colab. (May):1–12. Access date: 30 nov. 2020.
- [Blatt et al. 2017] Blatt, L., Becker, V., and Ferreira, A. (2017). Mapeamento Sistemático sobre Metodologias e Ferramentas de apoio para o Ensino de Programação. *Anais do XXIII Workshop de Informática na Escola (WIE 2017)*, 1(Cbie):815.
- [Carneiro et al. 2018] Carneiro, T., Medeiros Da Nóbrega, R. V., Nepomuceno, T., Bian, G., De Albuquerque, V. H. C., and Filho, P. P. R. (2018). Performance analysis of google colab as a tool for accelerating deep learning applications. *IEEE Access*, 6:61677–61685.
- [Davis et al. 1989] Davis, F. D., Bagozzi, R. P., and Warshaw, P. R. (1989). User acceptance of computer technology: a comparison of two theoretical models. *Management science*, 35(8):982–1003.
- [Delbem et al. 2019] Delbem, E., Motta, C., Oliveira, C., and Marques, C. V. M. (2019). Metodologia para analisar o desenvolvimento inter-relacional em alunos do ensino médio utilizando jogos. In *Anais Estendidos do XV Simpósio Brasileiro de Sistemas Colaborativos*, pages 160–161. SBC.
- [Farias et al. 2015] Farias, H., Bonifácio, B., and Ferreira, R. (2015). Avaliando o Uso da Ferramenta Scratch para Ensino de Programação através de Análise Quantitativa e Qualitativa. *Anais do XXVI Simpósio Brasileiro de Informática na Educação (SBIE 2015)*, 1(Sbie):947.
- [Grus 2019] Grus, J. (2019). *Data science from scratch: first principles with python*. O'Reilly Media.
- [Macnaughton 1996] Macnaughton, R. J. (1996). Numbers, scales, and qualitative research. *The Lancet*, 347(9008):1099–1100.
- [Mattos et al. 2017] Mattos, M., de Araújo, L., Zucco, F., Hein, N., da Cunha, K., de Oliveira, G., Bordone, L., Kruger, N., Schlögl, L., Tomé, A., et al. (2017). Aplicação da prática colaborativa no desenvolvimento de um jogo para o ensino da programação. In *Anais do XIV Simpósio Brasileiro de Sistemas Colaborativos*, pages 21–30. SBC.
- [Neto et al. 2018] Neto, A. S., Neto, F. M., Lima, R., Silva, S., and de Oliveira, E. J. (2018). Avaliação de um ambiente virtual gamificado para auxiliar o ensino-aprendizagem de estudantes de medicina. In *Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE)*, volume 29, page 496.

- [Neto and Schuvartz 2007] Neto, W. C. B. and Schuvartz, A. A. (2007). Ferramenta Computacional de Apoio ao Processo de Ensino-Aprendizagem dos Fundamentos de Programação de Computadores. *SBIE - Simpósio Brasileiro de Informática na Educação*, 17:520–528.
- [Vishakha Lall 2018] Vishakha Lall (2018). Google colab — the beginner’s guide. Access date: 1 jun. 2020.
- [Wainerman 1976] Wainerman, C. (1976). *Escalas de medición en ciencias sociales*. Ediciones Nueva Visión.
- [Wilder 1978] Wilder, J. W. (1978). *New concepts in technical trading systems*. Trend Research.