

Literacia Digital: Aprendendo a partir de dados

Ana Carolina R. Prates, Mauricio J. Bastos, Carlos Eduardo S. Silva,
Anderson C. de Lima¹, Valéria Q. Reis^{1,2}

¹Universidade Federal do Mato Grosso do Sul
Campo Grande, Brasil

²Leuphana Universität Lüneburg
Lüneburg, Alemanha

{ana.prates, m.jornada, carlos.e.s.silva}@ufms.br,

{anderson.lima, valeria.reis}@ufms.br,

Abstract. *In the context of an increasingly digitalized society, the ability to extract and interpret information from data has become essential for all citizens, especially educators. Primary and secondary school teachers play a crucial role in shaping individuals capable of developing the skills required in the digital age. In response to this need, this proposal presents an introductory Data Science course designed to equip teachers with analytical and technological approaches applied to education. The teaching materials are based on two pioneering projects widely recognized in the international academic community, ensuring an innovative approach aligned with global best practices.*

Resumo. *No contexto de uma sociedade cada vez mais digitalizada, a capacidade de extrair e interpretar informações a partir de dados tornou-se essencial para todos os cidadãos, especialmente para educadores. Professores da educação básica desempenham um papel crucial na formação de indivíduos aptos a desenvolver as competências exigidas na era digital. Diante dessa necessidade, esta proposta apresenta um curso introdutório de Ciência de Dados, voltado para capacitar docentes no uso de abordagens analíticas e tecnológicas aplicadas à educação. O material didático é fundamentado em dois projetos pioneiros amplamente reconhecidos na comunidade acadêmica internacional, garantindo uma abordagem inovadora e alinhada às melhores práticas globais.*

1. Descrição Geral

A Ciência de Dados consiste no estudo de dados a fim de extrair informações relevantes nos mais diversos domínios. Ela envolve a coleta, processamento, análise e interpretação de grandes volumes de dados para tomada de decisão [Provost and Fawcett 2013]. Dessa maneira, a Ciência de Dados ajuda no entendimento do porquê determinados fenômenos ocorrem, quais suas consequências e o que pode ser feito a respeito.

Aprender a partir de dados é uma habilidade que envolve pensamento computacional, mineração de dados, programação, estatística e aprendizado de máquina. Por consequência, ela é fortemente trabalhada em cursos de Computação. Porém, todos os tipos de profissionais podem se beneficiar da capacidade de extrair informações a partir de dados e, além disso, com os conhecimentos de sua área de atuação, podem colaborar significativamente para a interpretação de fenômenos.

No Brasil, a Base Nacional Comum Curricular (BNCC) prevê o ensino de conceitos relacionados à Ciência de Dados nos ensinos Fundamental e Médio, especialmente dentro da área de Matemática e Pensamento Computacional [Ministério da Educação (MEC) 2018, Ministério da Educação (MEC) 2022]. Dessa maneira, é importante que professores da Educação Básica sejam capacitados nessa temática a fim de interpretar gráficos, tabelas e estatísticas corretamente. Isso os ajuda na formação de jovens e adultos com pensamento crítico e capacitados para as profissões do futuro.

Baseado nas experiências de democratização do ensino de Ciência de Dados da Universidade da Califórnia (UC) – chamado Data 8 – e da Leuphana Universität Lüneburg (LUL) – denominado DATAx –, esta proposta prevê a expansão do conhecimento da análise de dados com o uso de ferramentas de programação para a comunidade de professores ativos no ensino médio e também professores em formação [Data 8, DATAx]. O material de ensino foi inicialmente elaborado e validado por membros de um projeto de ensino da Faculdade de Computação da Universidade Federal de Mato Grosso do Sul (UFMS) e será oferecido por meio de uma ação de extensão já devidamente cadastrada com a instituição acadêmica.

O curso será ofertado totalmente a distância, com material disponibilizado via Ambiente Virtual de Aprendizagem (AVA). Quatro acadêmicos da Faculdade de Computação da UFMS já estão capacitados para atuar como tutores.

2. Objetivos

Os objetivos gerais da proposta são:

1. Democratizar o acesso ao ensino da Computação;
2. Capacitar professores com metodologias exploratórias para o aprendizado a partir de dados;
3. Emponderar professores no entendimento de fenômenos dentro da sua área de conhecimento.

Os objetivos específicos consistem em:

1. Ensinar conceitos básicos de Estatística, Programação e Aprendizado de Máquina para a observação de fenômenos;
2. Apresentar aos docentes ferramentas que possibilitem a análise de grandes volumes de dados;
3. Exercitar através de casos reais o entendimento de fenômenos a partir de dados;
4. Capacitar profissionais para empregar Ciência de Dados em seus trabalhos.

3. Público-Alvo

O público-alvo compreenderá até 15 participantes, entre professores em formação ou já atuantes no ensino básico. Cada tutor do curso acompanhará 5 participantes. Não há pré-requisitos para a participação.

4. Habilidades Exploradas

O Pensamento Computacional é intensamente explorado durante o curso. Ele permite que os participantes questionem e avaliem, de maneira determinada, reflexiva e responsável, declarações e contextos da era digital.

O curso ensina conceitos e habilidades essenciais em programação de computadores e inferência estatística, em conjunto com a análise prática de conjuntos de dados do mundo real, incluindo dados econômicos, coleções de documentos, dados geográficos e redes sociais. Ele também aborda questões sociais relacionadas à análise de dados, como privacidade e design.

Mais especificamente, busca-se enriquecer o conhecimento dos educadores a fim de explorar parcial ou integralmente as seguintes habilidades da BNCC:

- (EM13CO12) Produzir, analisar, gerir e compartilhar informações a partir de dados, utilizando princípios de ciência de dados.
- (EM13MAT202) Planejar e executar pesquisa amostral sobre questões relevantes, usando dados coletados diretamente ou em diferentes fontes, e comunicar os resultados por meio de relatório contendo gráficos e interpretação das medidas de tendência central e das medidas de dispersão (amplitude e desvio padrão), utilizando ou não recursos tecnológicos.
- (EM13MAT406) Construir e interpretar tabelas e gráficos de frequências com base em dados obtidos em pesquisas por amostras estatísticas, incluindo ou não o uso de softwares que inter-relacionem estatística, geometria e álgebra.
- (EM13MAT407) Interpretar e comparar conjuntos de dados estatísticos por meio de diferentes diagramas e gráficos (histograma, de caixa (box-plot), de ramos e folhas, entre outros), reconhecendo os mais eficientes para sua análise.
- (EM13MAT510) Investigar conjuntos de dados relativos ao comportamento de duas variáveis numéricas, usando ou não tecnologias da informação, e, quando apropriado, levar em conta a variação e utilizar uma reta para descrever a relação observada.
- (EM13CO10) Conhecer os fundamentos da Inteligência Artificial, comparando-a com a inteligência humana, analisando suas potencialidades, riscos e limites.
- (EM13CO21) Comunicar ideias complexas de forma clara por meio de objetos digitais como mapas conceituais, infográficos, hipertextos e outros.

5. Recursos e Materiais Utilizados

O material a ser utilizado está disponível no endereço https://github.com/valeriaquadros/Data8_UFMS e já foi validado por cinco pessoas do meio acadêmico. Trata-se da tradução do curso *Data 8* ofertado no ano de 2022 com novas formulações de exercícios para serem utilizados como avaliações no curso proposto [Data 8]¹. A adaptação do material também buscou integrar parte da metodologia adotada no projeto DATAx [DATAx].

O Projeto *Data 8: The Foundations of Data Science* foi criado em 2015 na Universidade da Califórnia e tem sido ofertado três vezes ao ano a todos os acadêmicos daquela instituição. O material está disponível na Internet sob licença Creative Commons. A UC oferece suporte técnico para a replicação do projeto em instituições ao redor do mundo. Não é de nosso conhecimento a oferta do curso na língua portuguesa.

O Projeto DATAx tem sido implementado na Universidade Leuphana desde 2018 como forma de instigar acadêmicos de todos os cursos recém-ingressados na instituição.

¹<https://github.com/data-8/materials-sp22>

O Moodle da UFMS será utilizado como AVA e hospedará o material de ensino, assim como concentrará as avaliações do curso. Parte dessas avaliações consistirá em atividades auto-corrigíveis via plugin code runner. Cada participante será responsável por garantir computador próprio e acesso à Internet para participar da capacitação.

Três tutores, acadêmicos de Cursos de Computação da UFMS, acompanharão o aprendizado dos participantes, tirando dúvidas e avaliando as atividades entregues. Dois docentes da Faculdade de Computação, autores desta proposta, também acompanharão o desenvolvimento do curso.

6. Metodologia Detalhada de Desenvolvimento da Atividade

A oferta do curso está dividida em 4 fases: preparação, abertura das atividades, aplicação do curso e fechamento.

6.1. Preparação

Antes de iniciar o curso, os participantes deverão realizar as inscrições por meio de um formulário. Os usuários e senhas serão gerados pelo administrador do sistema para os cursistas inscritos.

6.2. Abertura da atividade

As atividades serão iniciadas com uma sessão de abertura para os alunos, onde será explicado o plano de ensino do curso. Durante essa sessão, o docente apresentará de forma geral os conteúdos e desafios abordados ao longo do curso, assim como o formato e datas das avaliações.

Na ocasião, será solicitado que cada participante preencha um questionário a fim de obter informações a respeito de sua área de formação, grau de conhecimento sobre o assunto, expectativas quanto ao curso e dados socio-demográficos.

6.3. Aplicação do curso

No decorrer do curso, os participantes poderão tirar suas dúvidas diretamente com os tutores uma vez por semana em horário pré-definido ou sob demanda. Estão previstas 13 atividades de avaliação que englobam os 15 módulos do curso. O período de duração do curso será de 5 semanas.

O programa do curso consiste nos seguintes tópicos:

1. Ciência de Dados
2. A causalidade e os experimentos
3. Programação em Python
4. Tipos de dados
5. Sequências de Sequências
6. Tabelas de mesas
7. Visualização
8. Funções e Tabelas
9. Aleatoriedade
10. Amostragem e distribuições empíricas
11. Teste de hipóteses

12. Comparando duas amostras
13. Estimativa
14. Por que o meio importa
15. Previsão
16. Inferência para a regressão
17. Classificação
18. Atualizando as previsões

6.4. Fechamento

Durante o encerramento, o docente instrutor solicitará que os participantes compartilhem suas experiências, destaquem os desafios enfrentados e os sucessos alcançados. O docente pode facilitar uma discussão sobre como os conceitos aprendidos podem ser aplicados em contextos particulares de cada cursista. Por fim, os participantes responderão a um questionário para avaliar o curso ministrado. Este questionário é de suma importância, pois fornecerá informações valiosas para a melhoria contínua do material e didática de ensino.

7. Avaliação

O curso terá como avaliação a entrega de 13 atividades. O cumprimento de 75% delas é necessário para a certificação do participante. Além disso, uma nota média igual ou superior a 6,0 nas avaliações é necessária para o aproveitamento das 60 horas de curso.

Considerando que esta proposta consiste em uma atividade de ensino, é primordial que uma análise seja realizada após o término da ação. Dessa maneira, se relata a experiência obtida, destacando tanto pontos de sucesso no projeto quanto quesitos que podem ser melhorados.

Além da avaliação quantitativa dos participantes por meio de provas, será aplicado, ao final das atividades, um questionário para uma análise qualitativa da satisfação dos educadores.

Referências

- Data 8. UC Berkeley Data 8. Acessado em 5 de fevereiro de 2025. Disponível em <https://www.data8.org/>.
- DATAx. DATAx — Leuphana. Acessado em 5 de fevereiro de 2025. Disponível em <https://www.leuphana.de/lehre/lehrentwicklung/projekte/abgeschlossene-projekte/datax.html>.
- Ministério da Educação (MEC) (2018). Base Nacional Comum Curricular (BNCC). Acessado em 5 de fevereiro de 2025.
- Ministério da Educação (MEC) (2022). Computação na Educação Básica - Complemento à BNCC. Acessado em 10 de fevereiro de 2025.
- Provost, F. and Fawcett, T. (2013). Data Science and its Relationship to Big Data and Data-Driven Decision Making. *Big Data*, 1(1):51–59.