

Data Science Workflow Canvas*

Start here. The sections below are ordered intentionally to make you state your goals first, followed by steps to achieve those goals. You're allowed to switch orders of these steps!

Title:		
<div>1 Problem Statement What problem are you trying to solve? What larger issues do the problem address?</div> <p>O problema proposto consiste na utilização de um modelo de aprendizado de máquina para a predição do IDHM com valores de produção agrícola de algumas culturas. Dessa forma é possível avaliar se a produção agrícola tem algum impacto no IDH em cada município brasileiro.</p>	<div>2 Outcomes/Predictions What prediction(s) are you trying to make? Identify applicable predictor (X) and/or target (y) variables.</div> <p>A variável alvo do dataset é: IDHM</p> <p>As variáveis de input são:</p> <ul style="list-style-type: none">-Prod_Soja-Prod_Café-Prod_Milho-Prod_Trigo-Prod_Algodao-Municipio	<div>3 Data Acquisition Where are you sourcing your data from? Is there enough data? Can you work with it?</div> <p>Os dados foram obtidos através de bases disponibilizadas nos sites:</p> <ul style="list-style-type: none">- http://www.atlasbrasil.org.br/ranking- https://www.inde.gov.br/VisualizadorMapas <p>O dataset final apresentou um total de 4.975 linhas.</p>
<div>4 Modeling What models are appropriate to use given your outcomes?</div> <p>Os modelos escolhidos foram: Decision Tree, KNN e SVM.</p> <p>Esses modelos foram escolhidos por suportarem problemas de classificação categórica.</p>	<div>5 Model Evaluation How can you evaluate your model's performance?</div> <p>Para avaliar a performance dos modelos, foi utilizado a biblioteca SKlearn, optando pelo classification report que retorna a Acurácia e precisão.</p> <p>No caso do KNN foi utilizado o Erro médio quadrático e o coeficiente de determinação.</p> <p>Ainda no caso do KNN, foi criado uma lista de valores binários indicando se a cada predição do modelo foi correta ou não.</p>	<div>6 Data Preparation What do you need to do to your data in order to run your model and achieve your outcomes?</div> <p>Para a aplicação do modelo, foi necessário a união dos dados de interesse em uma única planilha, a criação de uma coluna com dados categóricos e o tratamento dos valores nulos para 0.</p>

✓ Activation

When you finish filling out the canvas above, now you can begin implementing your data science workflow in roughly this order.

1 Problem Statement → 2 Data Acquisition → 3 Data Prep → 4 Modeling → 5 Outcomes/Preds → 6 Model Eval

* **Note:** This canvas is intended to be used as a starting point for your data science projects. Data science workflows are typically nonlinear.