

## Data Science Workflow Canvas \*

Titulo: Preditor de qualidade em vinhos

<b>1</b> <b>Problem Statement</b> <i>What problems are you trying to solve? What larger issues do the problem address?</i>	<b>2</b> <b>Outcomes/Predictions</b> <i>What predictions are you trying to make? Identify applicable predictor (x) and/or target (y) variables.</i>	<b>3</b> <b>Data Acquisition</b> <i>Where are you sourcing your data from? Is there enough data? Can you work with it?</i>
<p>Estamos explorando um conjunto de dados relacionado com as propriedades químicas de amostras de vinhos e tentando identificar quais são as características químicas que mais influenciam a nota de qualidade que um vinho irá receber e se é possível inferir qual avaliação de qualidade um vinho terá a partir de seus atributos físico-químicos.</p>	<p>Inicialmente identificaremos quais das variáveis possuem maior influência na nota de qualidade que é recebida e se essas várias são as mesmas para vinhos tintos e brancos.</p>	<p>Utilizaremos um dataset disponibilizado no repositório de Machine Learning da UCI, que conta com mais de 5.000 amostras de vinho verde, produzido em Portugal.</p>
<b>4</b> <b>Modeling</b> <i>What models are appropriate to use given your outcomes</i>	<b>5</b> <b>Model evaluation</b> <i>How can you evaluate your model's performance?</i>	<b>6</b> <b>Data preparation</b> <i>What do you need to do to your data in order to run your model and achieve your outcomes?</i>
<p>Exploraremos a criação de modelos voltados para os problemas de classificação, já que cada nota de qualidade pode ser interpretada como uma classe distinta.</p>	<p>Para avaliar a performance dos modelos compararemos os índices de acurácia obtida em cada um deles, fazendo uso da ferramenta Knime.</p>	<p>Como o dataset já havia sido utilizado em outros estudos acadêmicos não precisamos lidar com dados ausentes ou duplicados, ou mesmo com a descaracterização dos dados. Para alguns modelos criados foi necessária a normalização dos dados e outras parametrizações a fim de obter os melhores resultados.</p>

\* Adaptado de: <https://towardsdatascience.com/a-data-science-workflow-canvas-to-kickstart-your-projects-db62556be4d0>