**Dataset**

Dois conjuntos de dados -> amostras de vinho tinto e de vinho branco do “Vinho Verde” português.

**Entradas:** testes objetivos (valores de ph por exemplo). Variáveis físico-químicas.

**Saída:** dados sensoriais (mediana de pelo menos três avaliações feitas por especialistas em vinhos).

* Cada especialista avaliou a qualidade do vinho entre 0 (muito má) e 10 (muito bom).

Abordagem de **regressão** ou classificação.

As classes são ordenadas e não balanceadas. Por exemplo, há mais vinhos normais do que muito bons ou maus.

**Número de instâncias:**

* Vinho Tinto: 1599;
* Vinho Branco: 4898.

**Número de Atributos:** 11 + o atributo de saída.

Vários dos atributos podem ser correlacionados -> aplicar algum tipo de seleção de recurso.

**Variáveis de Entrada** (com base em testes físico-químicos):

1. Acidez fixa;
2. Acidez volátil;
3. Ácido cítrico;
4. Açúcar residual;
5. Cloretos;
6. Dióxido de enxofre livre;
7. Dióxido de enxofre total;
8. Densidade;
9. pH;
10. Sulfatos;
11. Álcool;

**Variável de saída** (com base em dados sensoriais):

1. Qualidade (pontuação entre 0 e 10).

pH afeta:

* aspeto visual;
* aroma;
* paladar;
* longevidade de um vinho.

A acidez vem da uva (ácido cítrico) e dá a sensação de frescor da bebida.