

O projeto desenvolvido irá prever a característica de um vinho se ele é tinto ou branco, com base em um dataset com as informações e detalhes de cada garrafa.

- Em um primeiro momento, foram criadas duas variáveis, sendo a variável “y” os dados da coluna “*style*”, que define se o vinho em questão é tinto ou branco, ou seja, os dados que iremos prever. E foi criada a variável “x” com as demais colunas do dataset, que são as variáveis preditoras.

```
In [ ]: # Separando as variáveis entre preditoras e variável alvo

y = arquivo['style']
x = arquivo.drop('style',axis = 1)
```

- No próximo passo, separei os dados em treino e teste, de maneira que fique evidente os dados utilizados na criação do modelo e aqueles utilizados apenas para testar seu desempenho.

Neste processo foi utilizado a função a *train_test_split()* do pacote sklearn. Assim garantimos que os dados serão aleatoriamente distribuídos entre os dois grupos e facilmente informamos o tamanho de cada grupo.

```
In [ ]: from sklearn.model_selection import train_test_split
#Criando o conjunto de dados do treino e teste

x_treino, x_teste, y_treino, y_teste = train_test_split(x, y, test_size = 0.3)
```

Machine Learning

Foi utilizado o algoritmo ExtraTrees, que criará várias árvores de decisão. Este algoritmo está pronto para ser utilizado, sendo necessário apenas indicar sua função. Como estamos trabalhando em um problema de classificação utilizaremos a função “ExtraTreesClassifier()”.

```
In [ ]: #Importando o algoritmo de árvore de decisão

from sklearn.ensemble import ExtraTreesClassifier

#Criação do modelo

modelo = ExtraTreesClassifier()
modelo.fit(x_treino,y_treino)

#Imprimindo resultados

resultado = modelo.score(x_teste,y_teste)
print("Acurácia:", resultado)
```

Com o método *"fit"* foi passado ao algoritmo as variáveis preditoras e a variável target, para que ele possa entender a relação entre estes dados e chegar ao modelo ideal. Com o método *"score"* indicamos ao modelo os dados de teste para que possamos avaliar seu desempenho.

O resultado final foi uma **acurácia superior a 99%**. Ou seja, em praticamente todos os testes realizados o modelo acertou sua previsão, indicando corretamente se determinado vinho é tinto ou branco.