

#Machine learning

In [42]:

```
#importando a bliblioteca pandas, e lendo o arquivo que tem os conjuntos de dados
#os dados são de características de um determinado vinho, e se ele é branco ou vermelho.
#o objetivo final é tentar prever qual a cor do vinho a parti de suas características usando
import pandas as pd
arquivo = pd.read_csv('C:\\Users\\thelo\\Analise de Dados\\Projetos\\DADOS\\wine_dataset.csv')
```

In [43]:

```
#ler as 5 primeiras linhas do conjunto de dados, e suas colunas.
arquivo.head()
```

Out[43]:

	fixed_acidity	volatile_acidity	citric_acid	residual_sugar	chlorides	free_sulfur_dioxide	total_sulfur_dioxide
0	7.4	0.70	0.00	1.9	0.076	11.0	14.5
1	7.8	0.88	0.00	2.6	0.098	25.0	16.0
2	7.8	0.76	0.04	2.3	0.092	15.0	14.5
3	11.2	0.28	0.56	1.9	0.075	17.0	14.5
4	7.4	0.70	0.00	1.9	0.076	11.0	14.5

In [44]:

```
#Muda o valor de red para 0 nas linhas, na coluna style // pra usar a sklearn os rotulos pr
# nesse caso como so tem dois tipos de vinho, decidi colocar 0 e 1
arquivo['style'] = arquivo['style'].replace('red',0)
```

In [45]:

```
##Muda o valor de white para 1 nas linhas, na coluna style
arquivo['style'] = arquivo['style'].replace('white',1)
```

In [46]:

```
#separa as variaveis
#y = coluna style, o rotulo, vermelho ou branco
y = arquivo['style']
# x = todas as colunas menos a style, as características, ph, qualidade etc...
x = arquivo.drop('style',axis = 1)
```

In [47]:

```
from sklearn.model_selection import train_test_split
#criando o conjunto de dados de treino e teste:
x_treino, x_teste, y_treino, y_teste = train_test_split(x,y, test_size = 0.3)
```

In [48]:

```
from sklearn.ensemble import ExtraTreesClassifier
#criando modelos
modelo = ExtraTreesClassifier()
modelo.fit(x_treino,y_treino)

#imPRIMir resultados
resultado = modelo.score(x_teste,y_teste)
print("acuracia", resultado)
```

acuracia 0.9953846153846154

In [49]:

```
#visualizando os dados a serem testados
y_teste[300:303]
```

Out[49]:

```
500      0
2333     1
1288     0
Name: style, dtype: int64
```

In [50]:

```
#visualizando os dados a serem testados
x_teste[300:303]
```

Out[50]:

	fixed_acidity	volatile_acidity	citric_acid	residual_sugar	chlorides	free_sulfur_dioxide	tc
500	7.8	0.52	0.25	1.9	0.081	14.0	
2333	6.9	0.38	0.25	9.8	0.040	28.0	
1288	7.0	0.60	0.30	4.5	0.068	20.0	

In [51]:

```
#Prevendo o resultado do rotulo, aparti das caracteristicas
previsao = modelo.predict(x_teste[300:303])
```

In [52]:

```
#printando a previsão
previsao
```

Out[52]:

```
array([0, 1, 0], dtype=int64)
```

