



MTRFoodDelivery vol ampliar la gamma de serveis i entrar en el profitós negoci vinícola. Com a expert en Intel·ligència Artificial obrirà una nova branca d'assistència a enòlegs, sommeliers i vinicultors.

Assessoraran sobre la qualitat del vi en funció de l'anàlisi dels següents paràmetres:

- fixed acidity
- volatile acidity

- citric acid
- residual sugar
- chlorides
- free sulfur dioxide
- total sulfur dioxide
- density
- pH
- sulphates
- alcohol

I s'especialitzaran en vins blancs i negres. Crearan un model diferent de Machine Learning per predir la qualitat de cada tipus de vi.

En el mercat han adquirit dos datasets de qualitat que els permetran entrenar per separat els dos models:

- winequality-red.csv
- winequality-white.csv

Sou contractats per la companyia com a Data Scientist per desenvolupar tots dos models.

El Data Scientist Chief us elabora el següent guió per facilitar-vos la tasca (preneu com a base l'script Kaggle_Titanic vist a classe de teoria):

- Carregueu els dos fitxers csv
- Transformeu la variable de resposta "quality" de la següent manera:
 - 0 (Vi de baixa qualitat): valors entre 0 i 4
 - 1 (Vi de qualitat mitja): valors 5 i 6
 - 2 (Vi de qualitat alta): valors entre 7 i 10
- Respongueu a través de gràfiques a les següents preguntes:
 - Al dataset de vi negre, quants vins hi ha de qualitat baixa, mitja i alta?
 - Al dataset de vi blanc, quants vins hi ha de qualitat baixa, mitja i alta?
 - Què caracteritza un vi negre de qualitat baixa? (quins valors mitjos té fixed acidity, volatile acidity, citric acid, residual sugar, chlorides, free sulfur dioxide, total sulfur dioxide, density, pH, sulphates i alcohol)

NOTA: utilitzeu el mètode *mean* de la llibreria *numpy*

- Què caracteritza un vi negre de qualitat mitja? (mateix criteri)
- Què caracteritza un vi negre de qualitat alta? (mateix criteri)
- Què caracteritza un vi blanc de qualitat baixa? (mateix criteri)
- Què caracteritza un vi blanc de qualitat mitja? (mateix criteri)
- Què caracteritza un vi blanc de qualitat alta? (mateix criteri)
- Dividiu els dos datasets en dades de train i test i guardeu el resultat en els nous datasets *trainNegre*, *testNegre*, *trainBlanc*, *testBlanc*.
NOTA: utilitzeu el mètode *train_test_split* de la llibreria *sklearn.model_selection* reservant un 70% per les dades de train i un 30% per les dades de test
- Afegiu un comentari al codi justificant la necessitat de dividir la mostra en train i test
- Comenta a més quina alternativa hi ha per obtenir el mateix objectiu?. Com funciona?
- Afegiu un comentari sobre quin mètode de Machine Learning utilitzareu argumentant la tria
- Pels dos models
 - Quin Accuracy heu obtingut?
 - Mostreu la Matriu de Confusió
 - Mostreu l'informe de Classificació per comprovar com de precís ha estat el mètode per cadascun dels tipus de qualitat
 - Justifiqueu si heu entrenat o no un bon model
- Afegeix un comentari definint què és i com es calculen l'Accuracy i la Matriu de Confusió

NOTA: utilitzeu la llibreria *sklearn.metrics*

NO HI HA TREBALL INDIVIDUAL PER AQUESTA ENTREGA

TREBALL EN GRUP (100%)

Implementa en python tots els passos indicats pel Data Scientist Chief. Utilitza els models entrenats per predir la qualitat de les següents mostres de vi:

NOTA: utilitza el mètode *predict* del model entrenat. Hauràs de convertir cadascuna de les mostres en un dataset que passaràs com a paràmetre al mètode *predict*

NOTA: per crear els datasets utilitza el mètode *DataFrame* de la llibreria *pandas*

- **Vi Negre:**
 - fixed acidity: 7.4
 - volatile acidity: 1.015
 - citric acid: 0.01
 - residual sugar: 2.9
 - chlorides: 0.075
 - free sulfur dioxide: 4
 - total sulfur dioxide: 11
 - density: 0.896
 - pH: 3.43
 - sulphates: 0.5
 - alcohol: 11.01

- **Vi Blanc:**
 - fixed acidity: 5.7
 - volatile acidity: 0.19
 - citric acid: 0.28
 - residual sugar: 10.6
 - chlorides: 0.028
 - free sulfur dioxide: 48
 - total sulfur dioxide: 134
 - density: 0.894
 - pH: 3.02
 - sulphates: 0.49
 - alcohol: 11.03