Project : Deep Learning

Subject : Predicting Wine Types: Red or White?

Deadline Monday December 5th 2022

For this project, you'll use the wine quality data set that you can find in the wine quality data set from the UCI Machine Learning Repository:

- https://archive.ics.uci.edu/ml/datasets/wine+Quality

The data consists of two datasets that are related to red and white variants of the Portuguese "Vinho Verde" wine. As stated in the description, you'll only find physicochemical and sensory variables included in this data set. The data description file lists the 12 variables that are included in the data, but for those who, like me, aren't really chemistry experts either, here's a short description of each variable:

1. *Fixed acidity*: acids are major wine properties and contribute greatly to the wine's taste. Usually, the total acidity is divided into two groups: the volatile acids and the nonvolatile or fixed acids. Among the fixed acids that you can find in wines are the following: tartaric, malic, citric, and succinic. This variable is expressed in g(tartaricacidtartaricacid)/dm3dm3 in the data sets.
2. *Volatile acidity*: the volatile acidity is basically the process of wine turning into vinegar. In the U.S, the legal limits of Volatile Acidity are 1.2 g/L for red table wine and 1.1 g/L for white table wine. In these data sets, the volatile acidity is expressed in g(aceticacidaceticacid)/dm3dm3.
3. *Citric acid* is one of the fixed acids that you'll find in wines. It's expressed in g/dm3dm3 in the two data sets.
4. *Residual sugar* typically refers to the sugar remaining after fermentation stops, or is stopped. It's expressed in g/dm3dm3 in the red and white data.
5. *Chlorides* can be a significant contributor to saltiness in wine. Here, you'll see that it's expressed in g(sodiumchloridesodiumchloride)/dm3dm3.
6. *Free sulfur dioxide*: the part of the sulfur dioxide that is added to a wine and that is lost into it is said to be bound, while the active part is said to be free. The winemaker will

always try to get the highest proportion of free sulfur to bind. This variable is expressed in mg/dm3dm3 in the data.

7. *Total sulfur dioxide* is the sum of the bound and the free sulfur dioxide (SO2). Here, it's expressed in mg/dm3dm3. There are legal limits for sulfur levels in wines: in the EU, red wines can only have 160mg/L, while white and rose wines can have about 210mg/L. Sweet wines are allowed to have 400mg/L. For the US, the legal limits are set at 350mg/L, and for Australia, this is 250mg/L.
8. *Density* is generally used as a measure of the conversion of sugar to alcohol. Here, it's expressed in g/cm3cm3.

9. *pH* or the potential of hydrogen is a numeric scale to specify the acidity or basicity the wine. As you might know, solutions with a pH less than 7 are acidic, while solutions with a pH greater than 7 are basic. With a pH of 7, pure water is neutral. Most wines have a pH between 2.9 and 3.9 and are therefore acidic.
10. *Sulfates* are to wine as gluten is to food. You might already know sulfites from the headaches that they can cause. They are a regular part of the winemaking around the world and are considered necessary. In this case, they are expressed in g(potassiumsulphatepotassiumsulphate)/dm3dm3.

11. *Alcohol*: wine is an alcoholic beverage, and as you know, the percentage of alcohol can vary from wine to wine. It shouldn't be surprised that this variable is included in the data sets, where it's expressed in % vol.

12. *Quality*: wine experts graded the wine quality between 0 (very bad) and 10 (very excellent). The eventual number is the median of at least three evaluations made by those same wine experts.

Project Steps: Deep Learning

1. Loading In the data,
2. Visualizing the data,
3. Train and Test Sets
4. Predict Values