Problem Statement - Wine Quality Prediction

Case Study: Diabetes Risk Prediction

Context:

Wine is a beverage made from fermented grapes and other fruit juices with a low amount of alcohol content. Wine is the second most popular alcoholic drink in the world after beer, and it is one of the most highly consumed beverages.

Generally, the quality of wine is graded based on the taste of the wine and vintage but this process is time-consuming, costly, and not efficient as the quality of the wine also depends on other physiochemical attributes like fixed acidity, volatile acidity, etc. Also, it is not always possible to ensure wine quality by experts when there is a huge demand for the product as it will increase the cost significantly.

Moonshine is a red wine company that produces premium high-quality wines. The company wants to improve its production efficiency and reduce the cost and additional time involved in wine tasting.

You as a data scientist at Moonshine company have to build a predictive model that can help to identify the premium quality wines using the available data.

Objective:

The data-set aims to answer the following key questions:

What are the different factors that can help in identifying a premium high-quality wine?

Can we build a model to identify high-quality wines? What should be the metric of choice to evaluate such a model?

Attribute Information:

- fixed acidity: Fixed Acidity impart sourness and resist microbial infection, measured in no. of grams of tartaric acid per dm3

- volatile acidity: No. of grams of acetic acid per dm3 of wine. Too high levels can

lead to an unpleasant, vinegar like taste

- citric acid: No. of grams of citric acid per dm3 of wine. Citric acid can add freshness and flavor to wines

- residual sugar: Remaining sugar after fermentation stops, measured in no. of grams per dm3.

- chlorides: No. of grams of sodium chloride i.e. salt per dm3 of wine

- free sulfur dioxide: No. of grams of free sulphite per dm3 of wine

- total sulfur dioxide: No. of grams of total sulphite (free sulphite+ bound) per dm3 of wine

- density: Density in gram per cm3

- pH: Describes how acidic or basic a wine is on a scale from 0 (very acidic) to 14 (very basic). The pH of the wine can help measure the ripeness of the wine

- sulphates: No. of grams of potassium sulphate per dm3 of wine

- alcohol: Volume of alcohol in percentage

- quality: Wine quality score between 3 to 8

Learning Outcomes:

- Exploratory Data Analysis
- Preparing the data to train a model
- Training and understanding of data using ensemble models
- Model evaluation

Steps and Tasks:

- Import Libraries and Load Dataset
- Overview of data
- Data Visualization
- Data preparation
- Choose Model, Train, and Evaluate
- Conclusion