Detailed Project Report (DPR)
Wine Quality Analysis

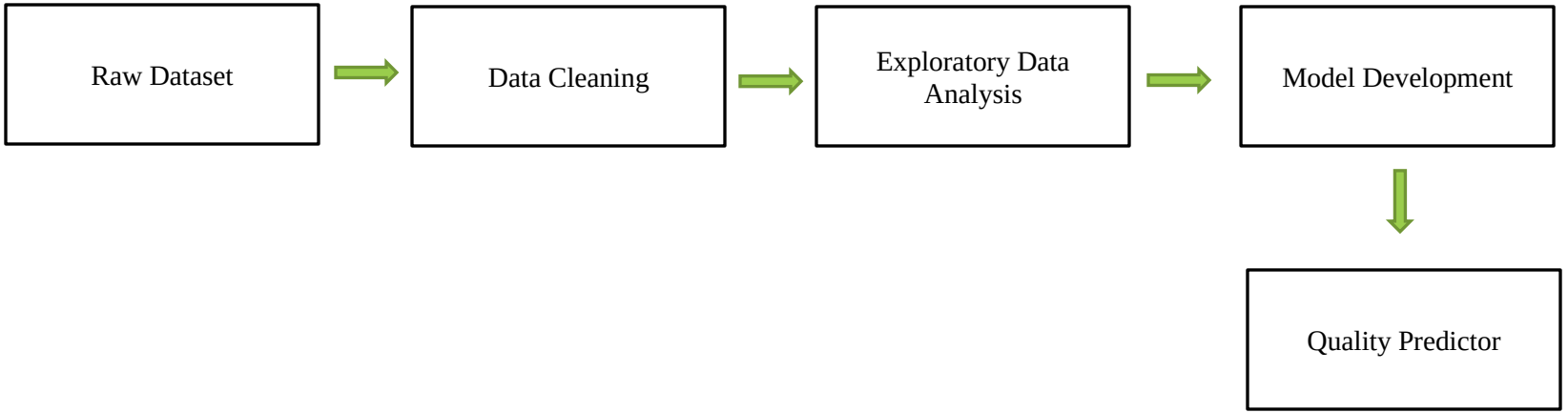
Document Version Control

Date Issue	Version	Description	Author
7/04/2023	1	Initial DPR - V 1.0	SHRUTI
14/04/2023	2	Initial DPR - V 2.0	SHRUTI

Objective

The objective of the project is to build a predictor that will help to understand the wine quality and its data pattern. The goal is to model wine quality based on physicochemical tests

Architecture



About the Project

There are Red and White Vinho Verde wine samples, from the north of Portugal. The goal is to model wine quality based on physicochemical tests.

This dataset contains details of the alcohol, residual sugar, acidity, density etc. and about having quality of wine. Using the attributes in the dataset we can predict quality of wine and identify the factors that affect the quality of wine.

Dataset Information

The two datasets are related to red and white variants of the Portuguese "Vinho Verde" wine. Due to privacy and logistic issues, only physicochemical (inputs) and sensory (the output) variables are available (e.g. there is no data about grape types, wine brand, wine selling price, etc.).

These datasets can be viewed as classification or regression tasks. The classes are ordered and not balanced (e.g. there are many more normal wines than excellent or poor ones). Outlier detection algorithms could be used to detect the few excellent or poor wines. Also, we are not sure if all input variables are relevant. So, it could be interesting to test feature selection methods.

Attribute Information

Input variables (based on physicochemical tests):

- fixed acidity
- volatile acidity
- citric acid
- residual sugar
- chlorides

Attribute Information

- free sulfur dioxide
- total sulfur dioxide
- density
- pH
- sulphates
- alcohol
- Output variable (based on sensory data):
- quality (score between 0 and 10)

Detailed Dataset Information

1. **Fixed acidity**: The concentration of non-volatile acids in wine, including tartaric, malic, and lactic acids. These acids affect the pH levels of the wine, which in turn can impact its taste and mouthfeel. For example, high levels of fixed acidity can make wine taste sour or tart, while low levels can make it taste flat.
2. **Volatile Acidity**: The concentration of acids, such as acetic acid, that can evaporate easily from wine and contribute to off-flavors and aromas. High levels of volatile acidity can give wine a vinegar-like taste and smell, while low levels can contribute to fruitiness and complexity.

3. **Citric acid** is a weak organic acid found naturally in citrus fruits. In winemaking, it can be added to adjust the acidity of the wine and enhance its flavor and aroma. Citric acid can also act as an antioxidant, helping to preserve the wine's color and freshness.
4. **Residual sugar** refers to the amount of sugar that remains in wine after fermentation. This can affect the wine's perceived sweetness, with higher levels of residual sugar making the wine taste sweeter.
5. **Chlorides** are salts, primarily sodium chloride, that can be found in wine. These salts can impact the taste and mouthfeel of the wine, with high levels of chlorides contributing to a salty or briny taste.

6. **Free sulfur dioxide** is a preservative commonly used in winemaking to prevent spoilage and oxidation. It can also act as an antioxidant, helping to maintain the wine's color and freshness.
7. **Total sulfur dioxide** refers to the total concentration of sulfur dioxide in wine, both free and bound. This measurement is important for assessing the wine's stability and safety, as high levels of sulfur dioxide can be harmful to human health.
8. **Density** is a measure of the mass per unit volume of wine. This measurement can provide information about the wine's alcohol and sugar content, as well as its overall viscosity and mouthfeel.

9. **pH** is a measure of the acidity or basicity of a solution, with a pH of 7 being neutral, lower values indicating acidity, and higher values indicating basicity. In winemaking, pH is an important factor that can affect the wine's taste, stability, and ability to age.
10. **Sulphates** are compounds that contain sulfur and oxygen. In winemaking, they can be added as a preservative or clarifying agent. While sulfates can help to prevent spoilage and improve the clarity of the wine, some people are sensitive to them and may experience adverse reactions.
11. **Alcohol** refers to the percentage of ethanol in wine, which affects its taste and mouthfeel. Higher levels of alcohol can contribute to a fuller body and a warm, burning sensation, while lower levels can make the wine taste lighter and more refreshing.

12. **Quality** refers to a score between 0 and 10 that is based on sensory evaluations of the wine's appearance, aroma, flavor, and overall appeal. The quality of wine can be influenced by a range of factors, including the grape variety, the winemaking process, and the environmental conditions in which the grapes were grown.

Wine Quality Prediction Web App

By SHRUTI KHANDELWAL

Quantity of Fixed Acidity

1.00

1.00 20.00

Quantity of Volatile Acidity

0.00

0.00 2.00

Quantity of Citric Acid

0.00

0.00 5.00

Quantity of Residual Sugar

0.00

0.00 20.00

Quantity of Chlorides

0.00

0.00 2.00

Quantity of Free Sulfur Dioxide

0.00

0.00 100.00

Quantity of Total Sulfur Dioxide

0.00

Quantity of Free Sulfur Dioxide

0.00

0.00 100.00

Quantity of Total Sulfur Dioxide

0.00

0.00 200.00

Density

0.00

0.00 1.00

Value of pH

0.00

0.00 14.00

Quantity of Sulphates

0.00

0.00 10.00

Quantity of alcohol

0.00

0.00 20.00

Wine Quality Results

Wine Quality is Bad

Q & A

Q1) What is the size of your data?

The size of the data in terms of KB is 381.0, with 6497 rows and 13 columns names.

Q2) What are the data type?

Initially, the columns consisted of all object, integer and float values. Later, object type was converted

Q & A

Q3) What were the libraries that you used Python?

The libraries were used are Pandas, NumPy, Matplotlib, Seaborn, and Streamlit.

Q4) What’s the complete flow you followed in this Project?

I had extracted the data into the Jupyter Notebook. After the essential steps of data cleaning and EDA, the model was developed using Random Forest Classifier for the prediction. Then it was deployed on Streamlit using GitHub repository.

Q & A

Q5) Where did you get the data ?

The data was provided by iNeuron.

https://drive.google.com/drive/folders/1FkmFVL8wlJmQWP1z52TD8PlhOJhitTyI?usp=sharing

THANK YOU