MACHINE LEARNING WINE QUALITY

MEMBERS

01

Nguyễn Trung Hậu

21110434

02

Mai Trần Bách

21110379

03

Huỳnh Hồng Khanh

21110499

04

Phạm Hữu Tuấn

21110713

CONTENTS

01. PURPOSE OF THE PROBLEM

- Built a machine learning model capable of predicting the quality of a type of wine.
- The model provides businesses with predictions of wine quality before testing, saving time and cost in the testing process.

02. DATA DESCRIPTION

- This datasets is related to red variants of the Portuguese "Vinho Verde" wine.The dataset describes the amount of various chemicals present in wine and their effect on it's quality.

DATA DESCRIPTION

FIXED ACIDITY

This is the fixed acidity level in wine. (g/L)

VOLATILE ACIDITY

This is the volatile acidity level in wine. (g/L)

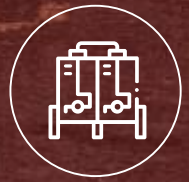
CITRIC ACID

Mức độ axit citric trong rượu. (g/L)

RESIDUAL SUGAR

This is the citric acid level in wine. (g/L)

DATA DESCRIPTION



CHLORIDES

The chloride level in wine. (g/L)

FREE SULFUR DIOXIDE

The amount of free sulfur dioxide in wine. (mg/L)

TOTAL SULFUR DIOXIDE

The citric acid level in the wine. (g/L)

DENSITY

The density of the wine. (g/cm^3)

DATA DESCRIPTION

PH

This is the pH level of the wine, measuring the acidity or alkalinity of the wine.

SULPHATES

Sulfate level in wine. (g/L)

ALCOHOL

Alcohol content of wine (%).

QUALITY

This is the target variable for the prediction model. (0-10)

METHODOLOGY USED IN THE PROBLEM

METHODOLOGY USED IN THE PROBLEM

Logistic Regression

$$f(x) = \frac{1}{(1 + e^{(-x)})}$$

Linear Regression

$$L_D(w) = \sum_{i=1}^{n} (w^T x^{(i)} - y^{(i)})^2$$

K-Nearest Neighbors

$$f(x) = \frac{1}{k} \sum_{x_i \in N_k(x,D)} y_i$$

- Using three models: Linear Regression, Logistic Regression, and k-NN to apply to the data and determine the best model for the dataset.

METHODOLOGY USED IN THE PROBLEM

Logistic Regression

- ✓ Easily understandable and implementable.
- ✓ Fast training speed.
- ✓ Can describe non-linear relationships.

Linear Regression

K-Nearest Neighbors

METHODOLOGY USED IN THE PROBLEM

Logistic Regression

Linear Regression

- ✓ Simple and easy to understand.
- ✓ Good for linear predictions.
- ✓ Fast training time.
- ✓ Strengths in regression.

K-Nearest Neighbors

METHODOLOGY USED IN THE PROBLEM

Logistic Regression

Linear Regression

K-Nearest Neighbors

- ✓ It can be used for various different problems.
- ✓ Suitable for large input variables.
- ✓ Easy to deploy and understand.

METHODOLOGY USED IN THE PROBLEM

Logistic Regression

↓

The wine quality problem involves classifying wines into different quality categories.

Linear Regression

↓

There is a linear relationship between the wine's characteristics and its quality.

K-Nearest Neighbors

↓

Wine quality can be influenced by many complex interacting factors.

EVALUATION OF EXPERIMENTAL RESULTS

The size of the dataset.

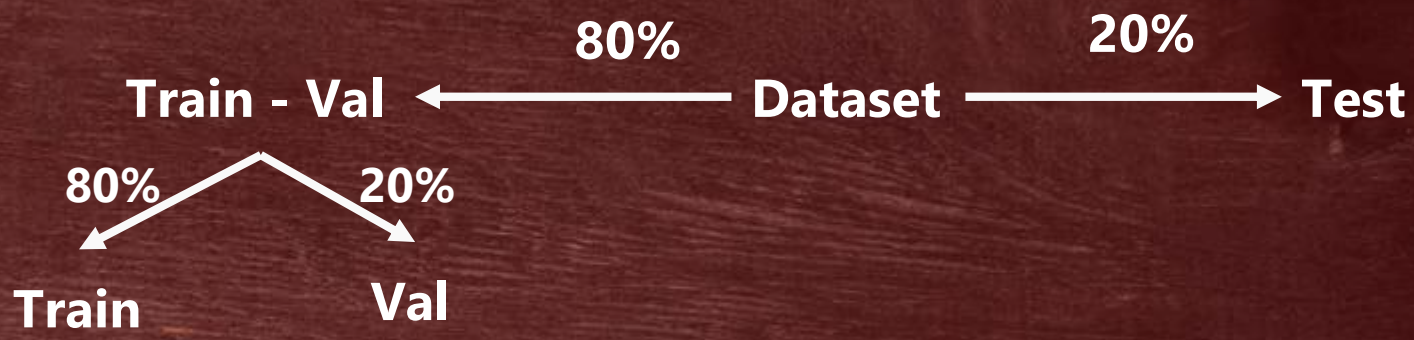
The dataset consists of 1143 rows and 13 attributes.

We need to build a model that should not be too complex to avoid overfitting.

EVALUATION OF EXPERIMENTAL RESULTS

Splitting the training and test ratios.

With 1143 rows of data, we will split it into two parts.



EVALUATION OF EXPERIMENTAL RESULTS

The error metrics used are:

Mean Square Error (MSE)

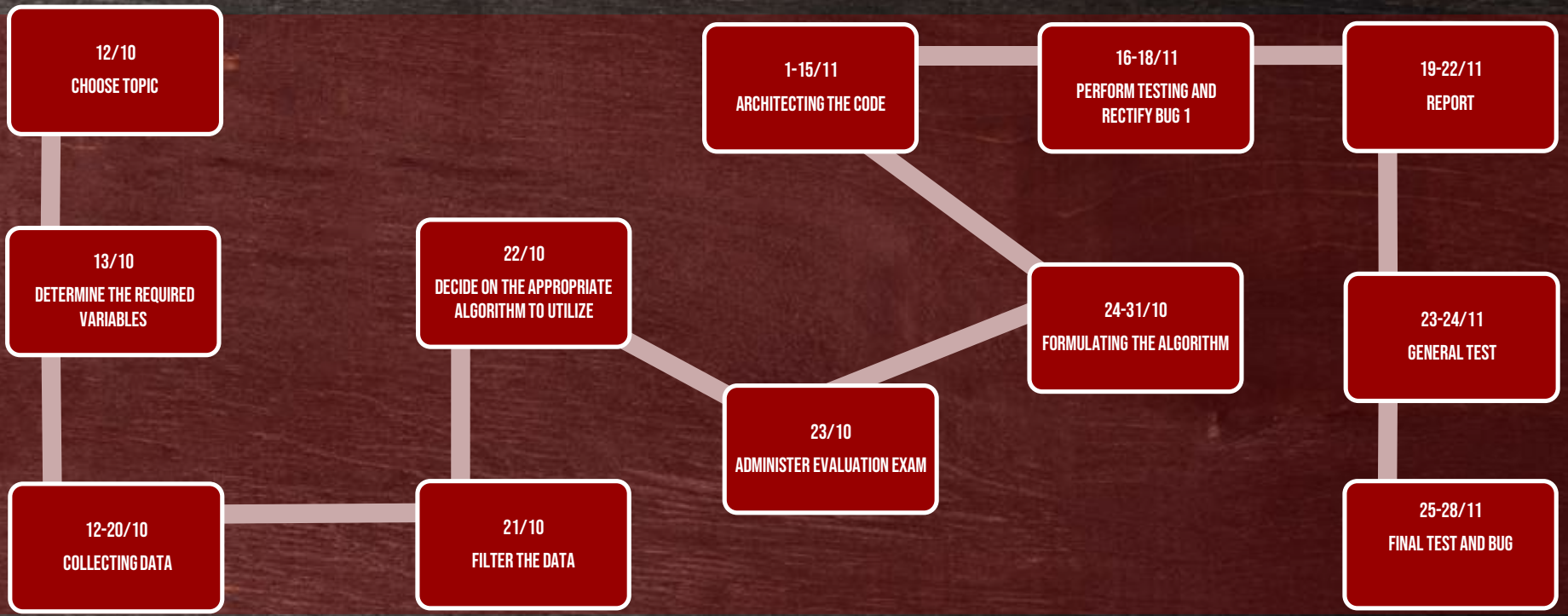
Mean Absolute Error (MAE)

Accuracy

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

$$\sum_{i=1}^{n}$$

WORK ASSIGNMENT PLAN



Thank you for listening