WINE QUALITY

AN UNSUPERVISED MACHINE LEARNING APPROACH

July 15th 2022

Mohamed MZAOUALI

OUTLINE

- Introduction
- Exploratory Data Analysis
- Feature Engineering
- Model Selection
- Data Analytics
- Conclusion
- References

INTRODUCTION

Objective:

- Clusteting
  - Quality level clusters
- Unsupervised Machine learning algorithms

Data Source:

- Wine quality data set

EDA

DATA SET:

Shape:

- 6497 Lines by 13 Columns.

Null Values:

- None.

Missing Values:

- None.

NaN Values:

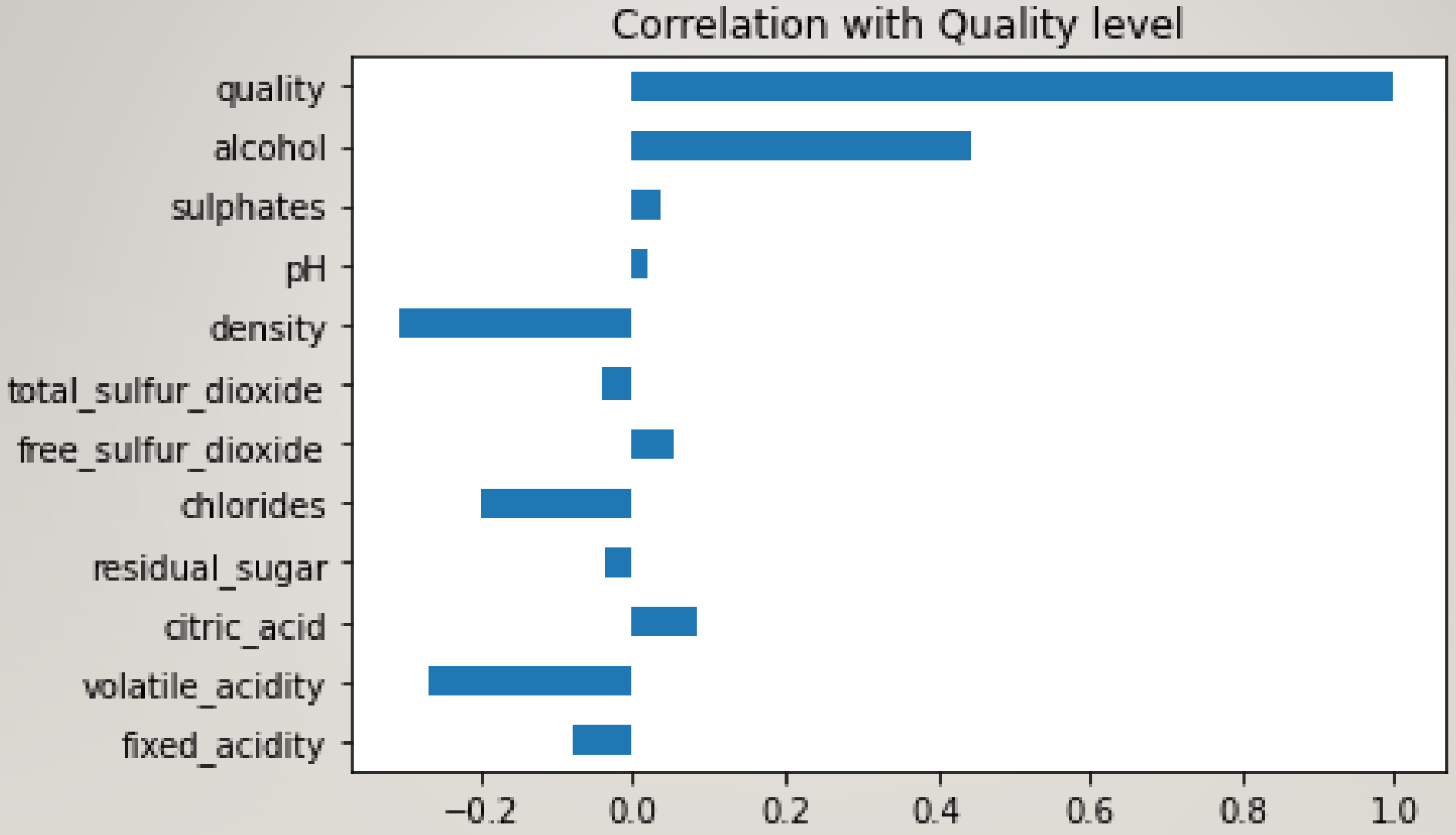
- None.

Independent Variables:

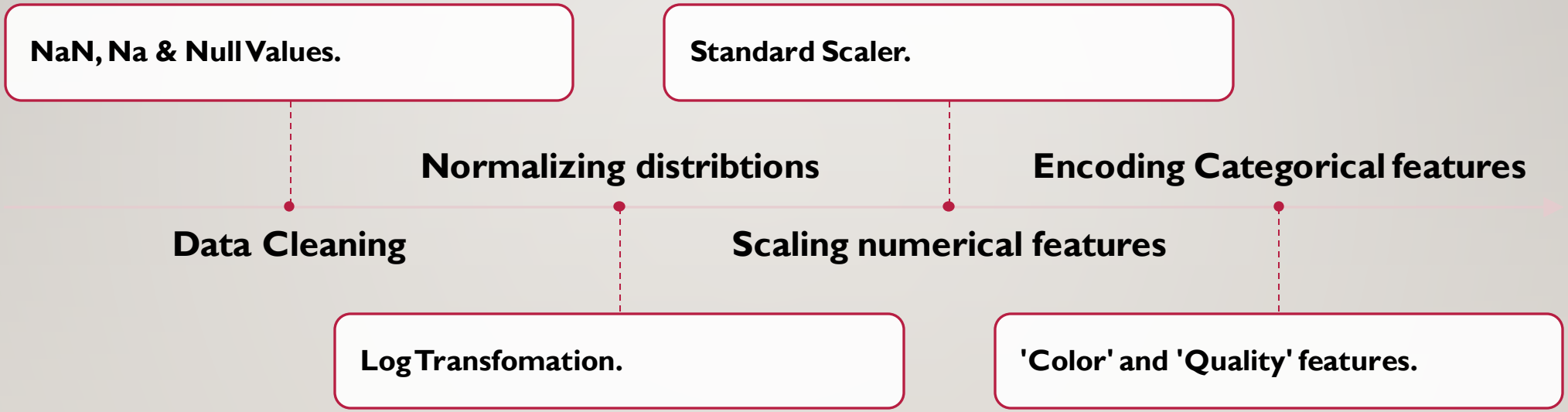
12 numerical features.

Dependent Variable:

'Quality' Column with levels from 3 to 9.


Correlation with Quality level
quality
alcohol
sulphates
pH
density
total_sulfur_dioxide
free_sulfur_dioxide
chlorides
residual_sugar
citric_acid
volatile_acidity
fixed_acidity
−0.2
0.0
0.2
0.4
0.6
0.8
1.0

FEATURE ENGINEERING



MODEL SELECTION

Clustering Models

K-means

Hierarchical Agglomerative Clustering

DBSCAN

Very low predictive power

Failure to recognize clusters

Results

DATA ANALYTICS

K-means	
Precision: 0.24	Recall: 0.06

Hierarchical Agglomerative Clustering	
Precision: 0.30	Recall: 0.14

Density-Based Spatial Clustering of Applications with Noise (DBSCAN)	
Clusters: 46	Noise points: 6196

CONCLUSION

- Insufficient data points &/or features
- Inadequate Approach

Failure to capture clusters:

Future Improvements:

- Multi-class regression approach
- Deep Learning Approach

REFERENCE

- The full approach is presented in the following ipynb:
- https://github.com/mzaoualim/Coursera_IBM_Machine_Learning_Professional_Certificate/blob/main/Unsupervised_Machine_Learning/Unsupervised_ML_Project.ipynb