

Machine Learning on Wine data set

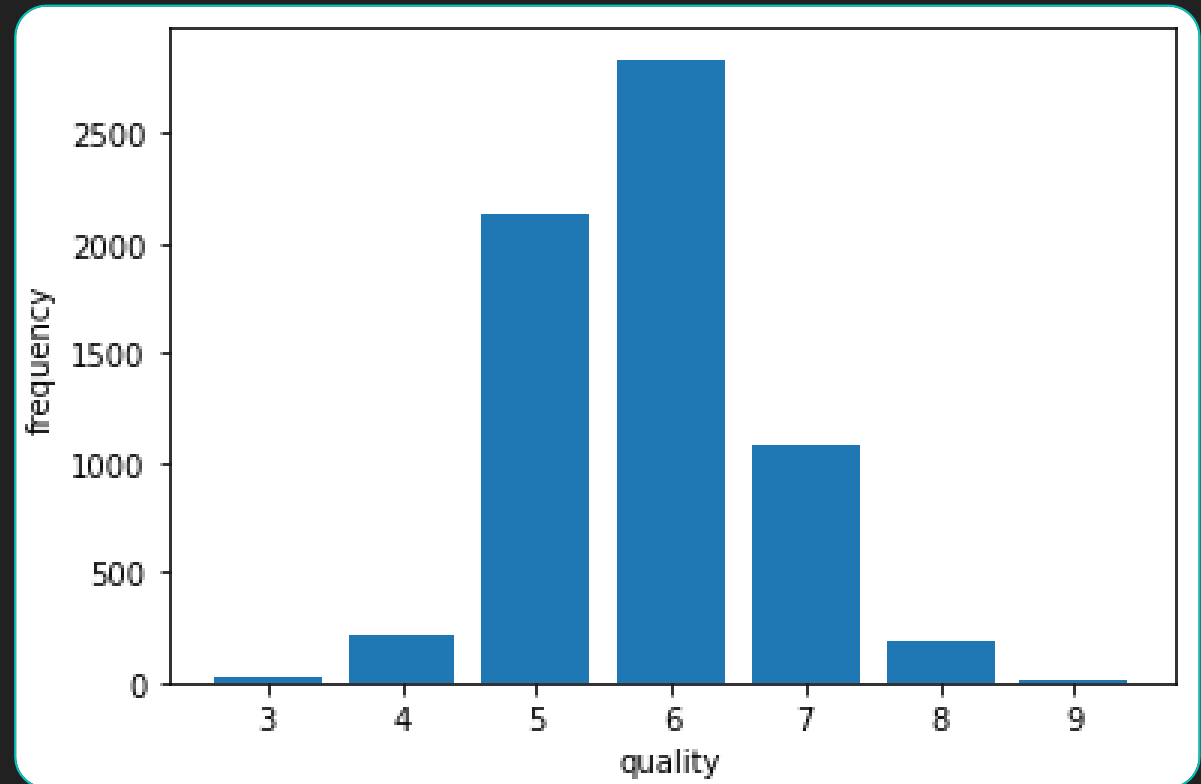
Elias Berglin, Gustav Nilsson



[Det här fotot](#) av Okänd författare licensieras enligt [CC BY-NC](#)

Data Understanding

- Examine data using info()
- Normalize Data
- Split Quality
 - Bad ≤ 5
 - Good > 5



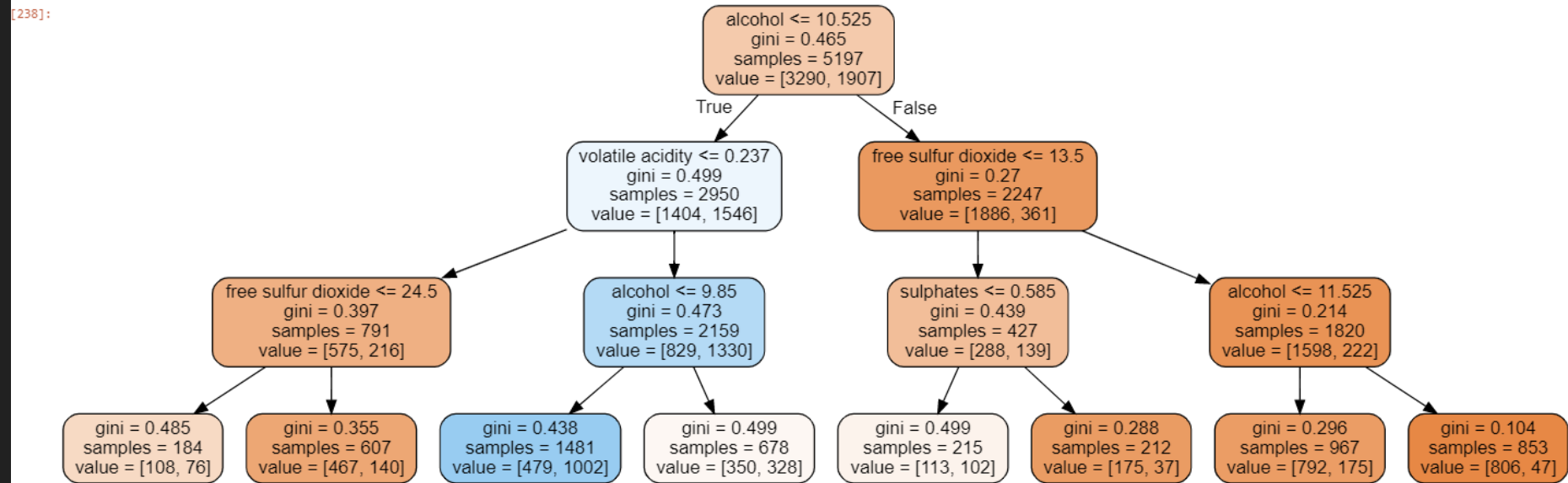
Data Preparation

- No null values
- Combine data set
- Split the data into a test and train set
- Normalize all the features values
- Change the target variable to “Good” or “Bad”

Models

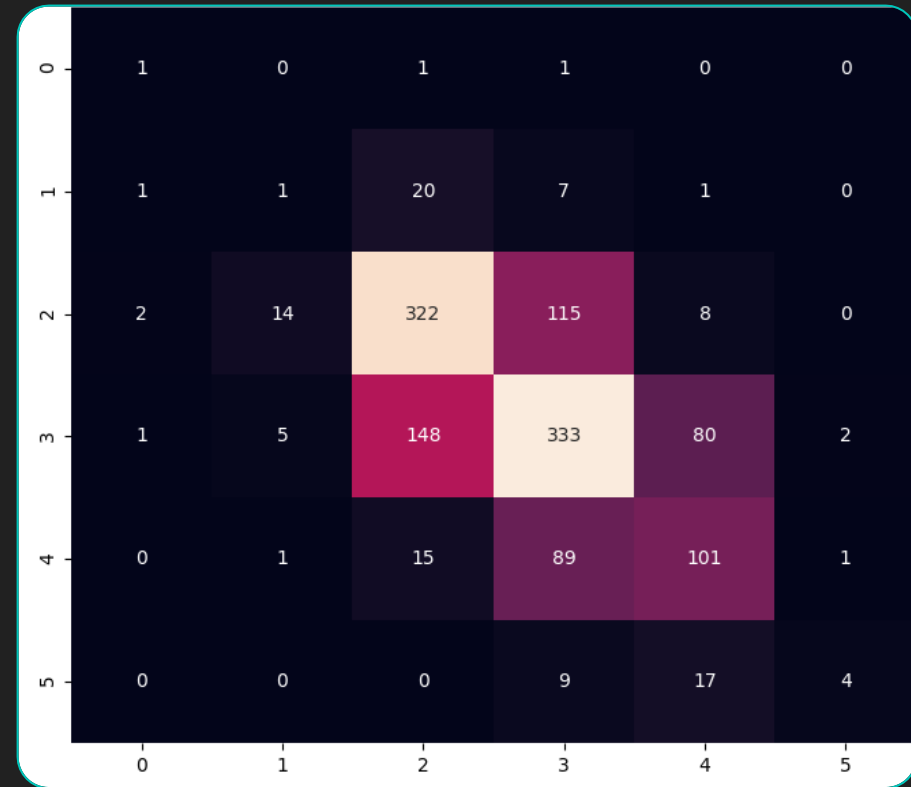
- ANN
 - Classification
 - Regression
- Decision Tree
 - Classification

```
model = keras.Sequential([  
    layers.Dense(256, activation='relu', input_dim=11),  
    layers.Dense(128, activation='relu'),  
    layers.Dense(1, activation='sigmoid')  
)  
model.compile(loss='binary_crossentropy', optimizer='adam', metrics=['accuracy'])
```



ANN – Regression

- MSE: 0.48
- Correct Guesses: 762
- Accuracy: 59%



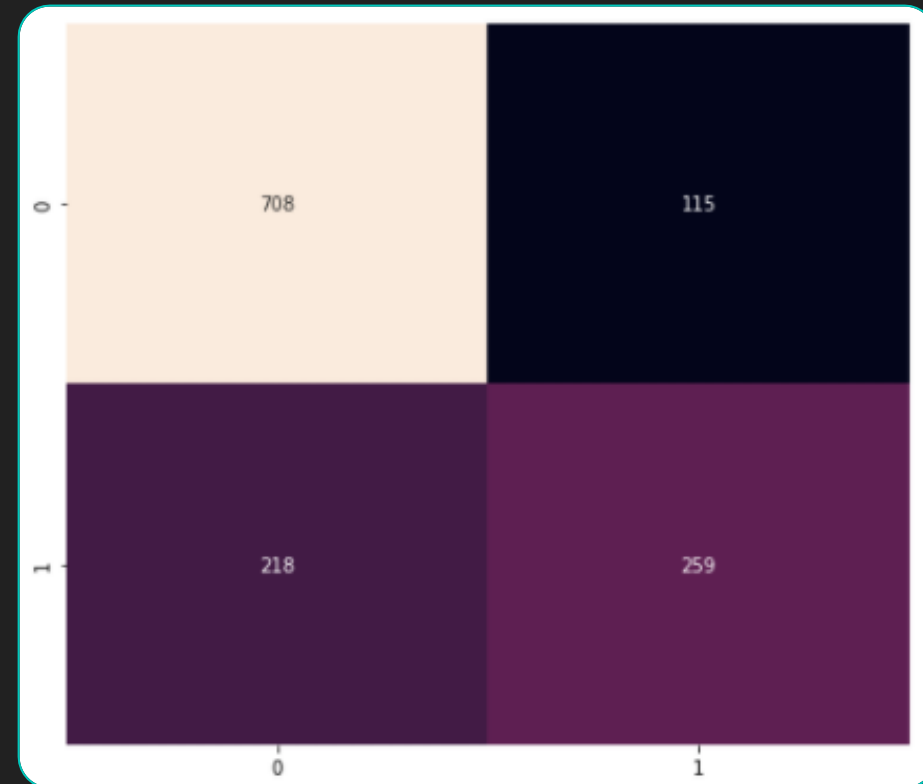
ANN – Classification

- Accuracy: 77%
- True Positive Rate: 72%
- True Negative Rate: 87%



Decision Tree – Classification

- Confusion Matrix
- Accuracy = 74%
- True Positive Rate: 69%
- True Negative Rate: 86%



Conclusion

- ANN Better Overall
- Regression
- Decision Tree was less consistent than ANN
- Alcohol most important feature for tree