

Predicting Wine Quality

Thirumurugan Vinayagam

Mohammed Saqib

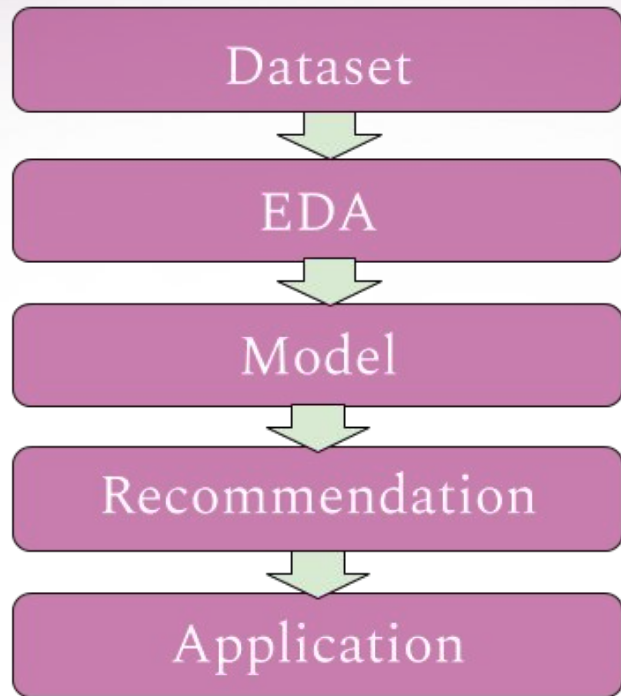
Shelby Watson

Amey Athaley

Daniel Oh



Our Process



Our problem and the dataset we used to form solutions

Exploratory Data Analysis and dataset observations

Trial of various supervised learning models and selection

Our model recommendation to best solve the problem

How our recommendations can be applied

What are we trying to predict?

- Wine Quality

- Ratings range from 3-9
- Median of 6
- Based on 12 predictors
 - Alcohol, density, pH, residual sugars, chlorides, citric acid, sulphates, free sulphur dioxide, total sulfur dioxide, volatile acidity

Dataset from UC Irvine Machine Learning Repository

Dataset

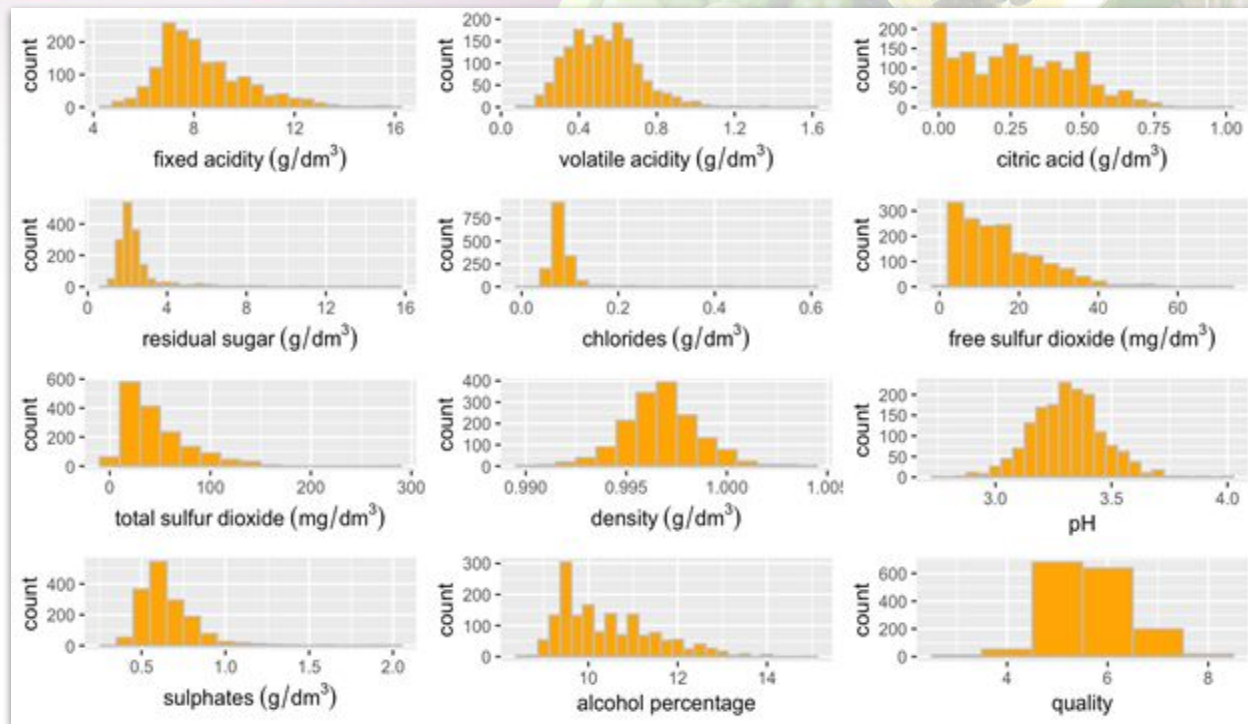
EDA

Model

Recommendation

Application

EDA



Dataset

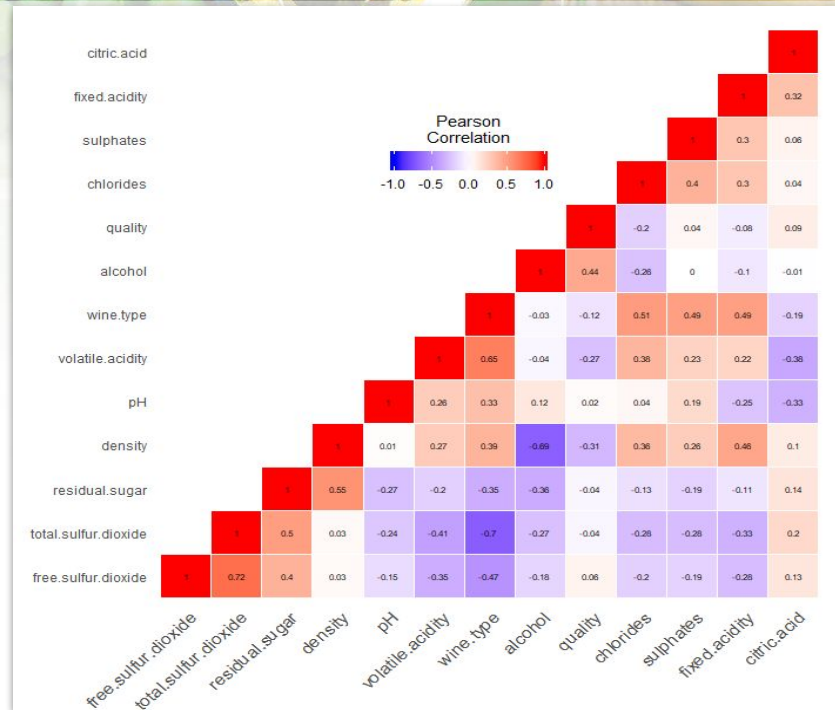
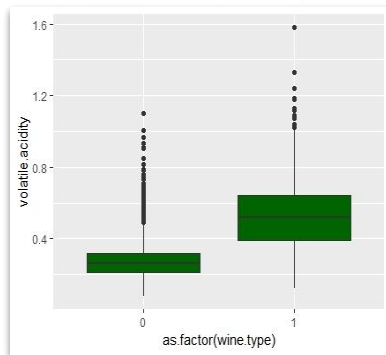
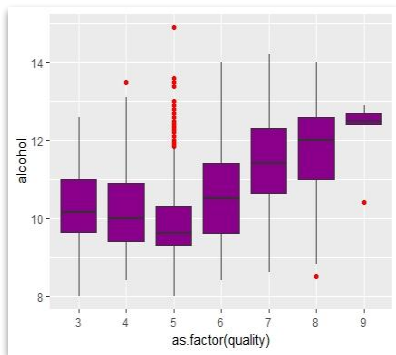
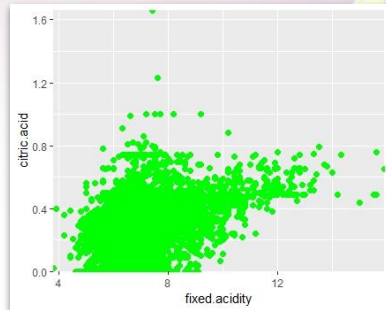
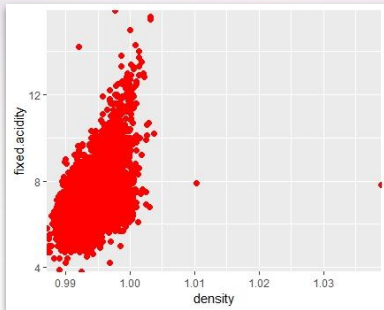
EDA

Model

Recommendation

Application

EDA



Dataset

EDA

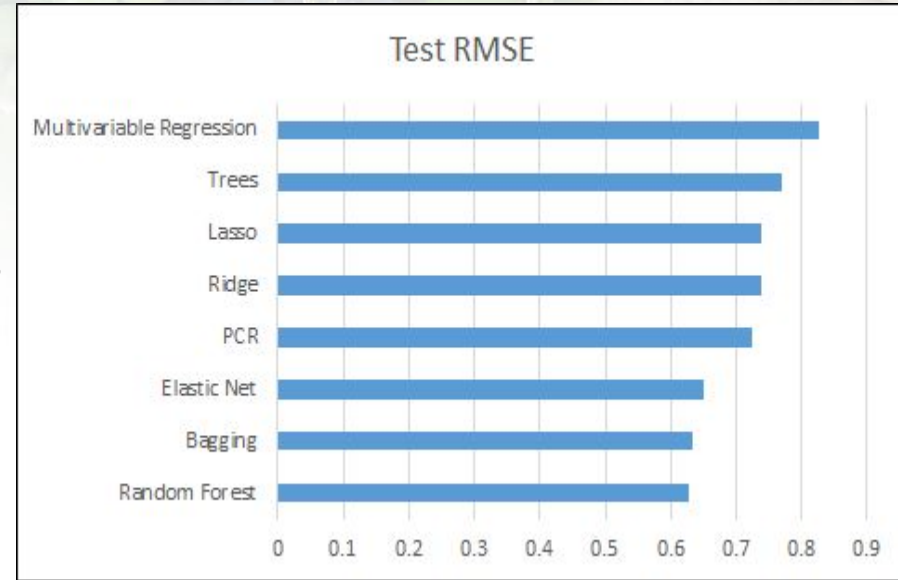
Model

Recommendation

Application

Model Results

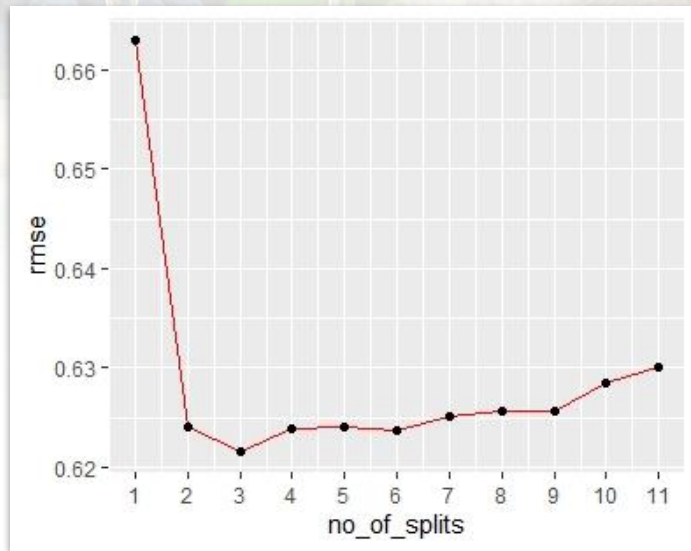
- Test RMSE:
 - Ridge: 0.737
 - Lasso: 0.738
 - Elastic Net: 0.649
 - Multi. Var Regression: 0.826
 - PCR: 0.723
 - Trees: 0.770
 - Bagging: 0.632
 - Random Forest: 0.623
 - Boosting: 0.699



Model Recommendation

We recommend the Random Forest algorithm:

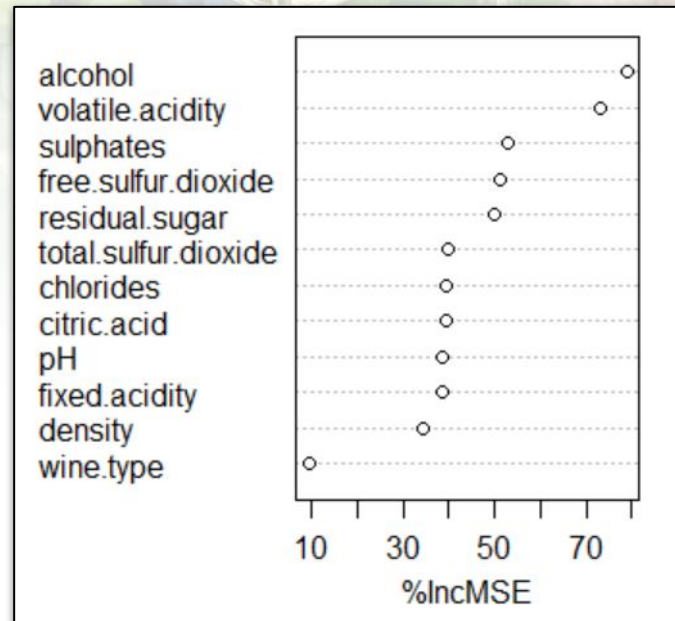
- Lowest RMSE of 0.623
- Quality Points
- ~46% Variance explained
 - Good model for subjective problem



Our Recommendation

Which variables are the most important?

- Alcohol Level
- Volatile Acidity



Application



Potential Clients

- Online Wine Retailers
- Grocery stores
- Wineries
- Restaurants



Dataset

EDA

Model

Recommendation

Application

Questions?

