**The Wine Land**

**Wine Data Analysis**

**Tanmay Jain**

[**tanmayjain84@gmail.com**](mailto:tanmayjain84@gmail.com)

**+91 9783466556**

**CONTENTS:**

1. Model used
2. Features extracted
3. Model accuracy

**Model Used:**

**Random Forest Classifier**

Random forests or random decision forests are an ensemble learning method for classification, regression and other tasks that operate by constructing a multitude of decision trees at training time and outputting the class that is the mode of the classes or mean prediction of the individual trees.

The most important parameter of the RandomForestClassifier class is the n\_estimators parameter. This parameter defines the number of trees in the random forest. We will start with n\_estimator=20 to see how our algorithm performs and see till n=50.

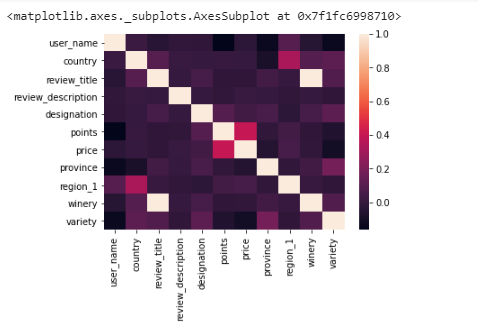
**Features Extracted:**

First the data is cleaned (Removing/filling Nan values with subsequent methods.)

Then data was encoded with the help of Ordinal Encoder

Correlation Matrix is calculated and the features are extracted on the basis of HeatMap.

**HEATMAP**



After analysing the heatmap, the p-value analysis was used to extract the features.

**MODEL ACCURACY:**

For classification problems the metrics used to evaluate an algorithm are accuracy, confusion matrix, precision recall, and F1 values. Execute the following script to find these values.

ACCURACY:

98.87%(at n\_estimator=20)

99%(at n\_estimator=50)



