Abstract:

Biannually, the Court of Master Sommeliers administer an exam to best of the best wine experts in the world. The candidates, who fail at a 94% rate, must have 10 years of sommeliers, and pass two qualifying exams, just to be able to take the masters’ test. This is a fittingly pretentious process to become one of histories’ 274 master sommeliers. But what about the rest of the world that doesn’t have over 10 years to dedicate to understanding wine. How is the average consumer supposed to navigate the wine menu? With data analysis, of course…

This project is meant to provide insight for the consumer when choosing which bottle of wine they want to buy. All wines are rated on a scale from 80-100, and have accompanying features such as country, category, vineyard, description, etc. The end goal is to use machine learning to build a model to classify bottles of wine, to know which bottles will lead to a positive critic score. Along the way, the project will provide insight into which features of wine yield the best price, score, and value ratio. I look at which country makes the most overpriced wine, which type of wine gives you the best bang for your buck, and which country is all price no points. The model uses the description of the bottle written by the critic to see what critics look for when determining if the bottle will get a positive of negative review. The methodology behind this is using NLP (Natural Language Processing) in order to analyze sentiment of review, and use this classification to predict whether the bottle’s rating will be above or below the mean score. I use Random Forest Classifier and a Recurrent Neural Network and compare the classification results between the two. This will not only help consumers get value from their wine selection, but also help producers know what qualities to focus on to get the best reviews.