
Sentiment Analysis For the Wine Connoisseur

Mark Ehler -

The Task at Hand

- Process the text of wine reviews to predict their score
- The problem - All scores are between 80-100 - all scores sound “good”

“Exaggeration is truth that has lost it’s temper.” - Khalil Gibran

- Personal taste - terroir - reviews are subjective - sponsorships
-

Methodology

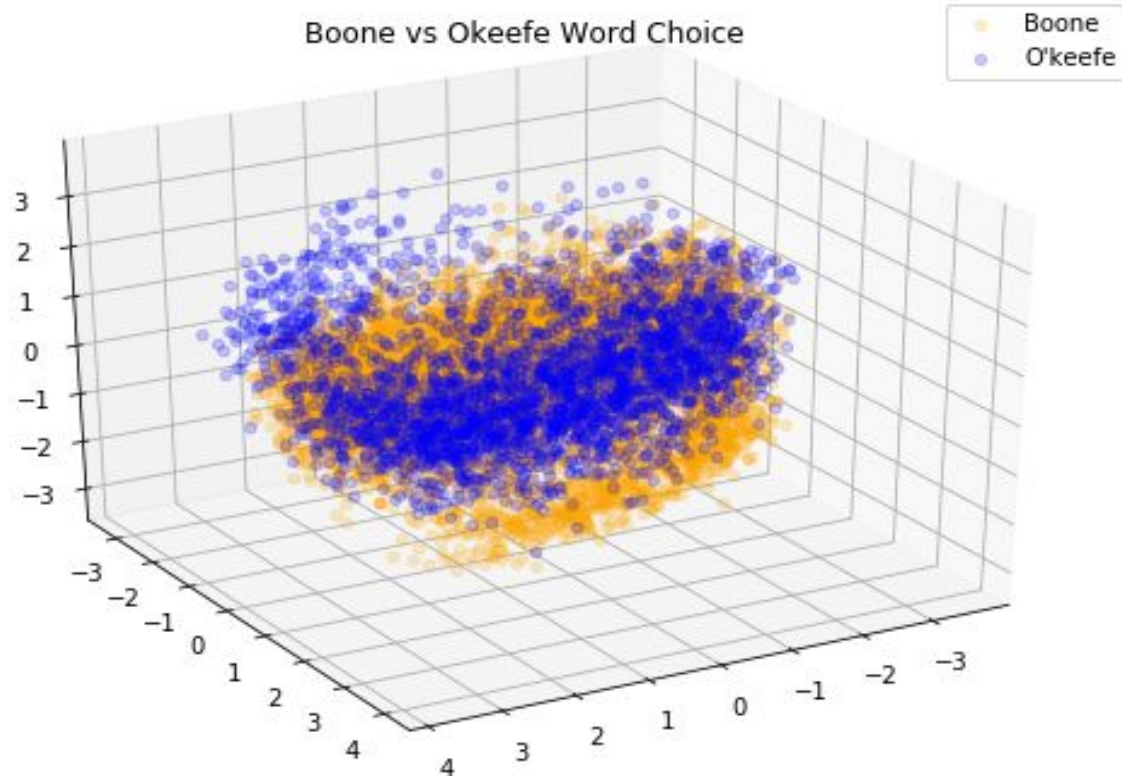
1 = The good scores 4 ● ● ● ● The bad scores = 0

Limited by only using the reviews of wines, will a deep learning neural network or a word2vec ensemble methods be better at predicting wine reviews?

Dataset - vectorized

All axes are representative of directional space.

Points close to each other should have related meaning.



Flavor notes

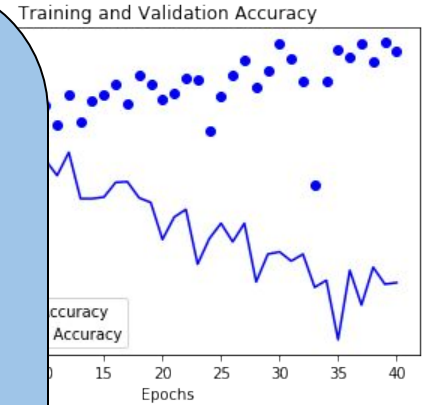
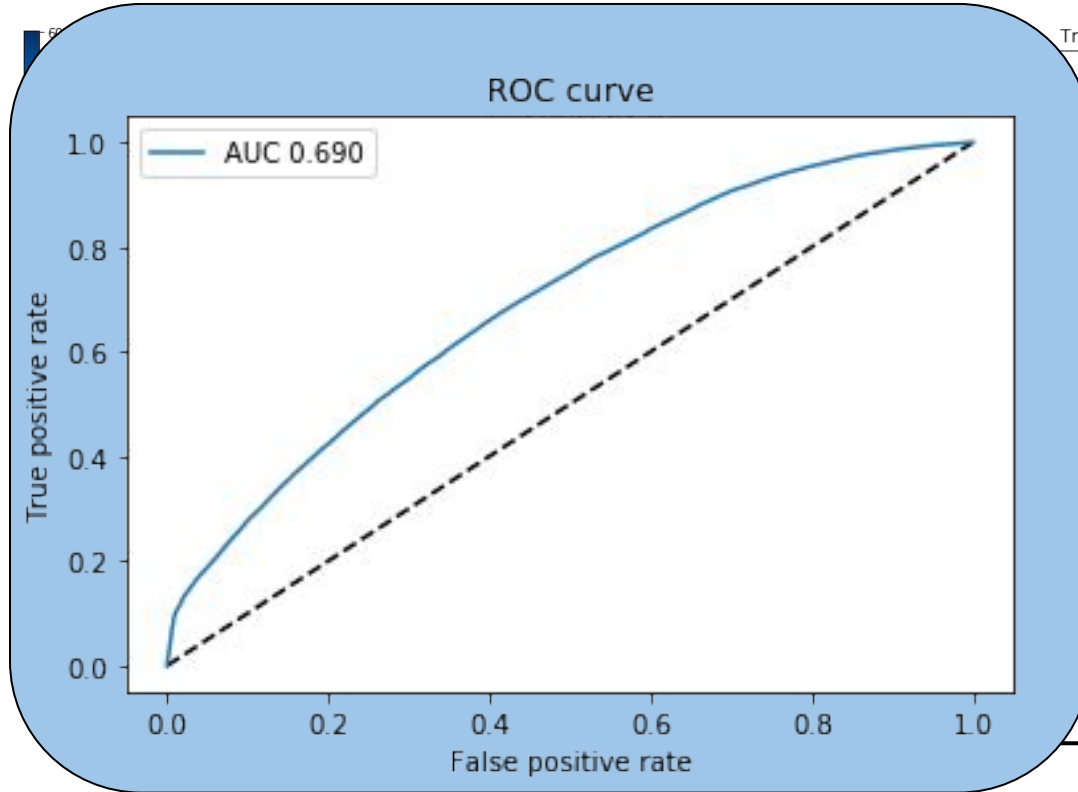
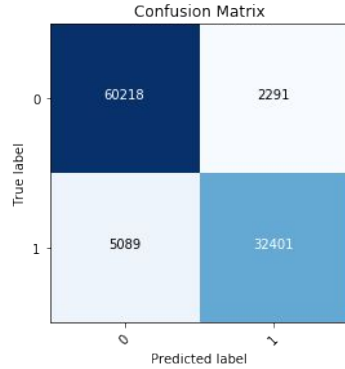
```
[23]: w2v_model.wv.most_similar('coffee')
```

```
[23]: [('espresso', 0.8969775438308716),  
      ('carob', 0.7564544677734375),  
      ('mocha', 0.7402285933494568),  
      ('licorice', 0.7393960952758789),  
      ('cocoa', 0.7026122212409973),  
      ('char', 0.6930364966392517),  
      ('cassis', 0.692969799041748),  
      ('molasses', 0.6914410591125488),  
      ('coconut', 0.673872709274292),  
      ('woodspice', 0.6711657047271729)]
```

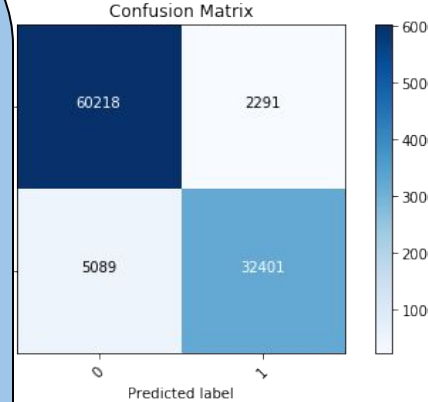
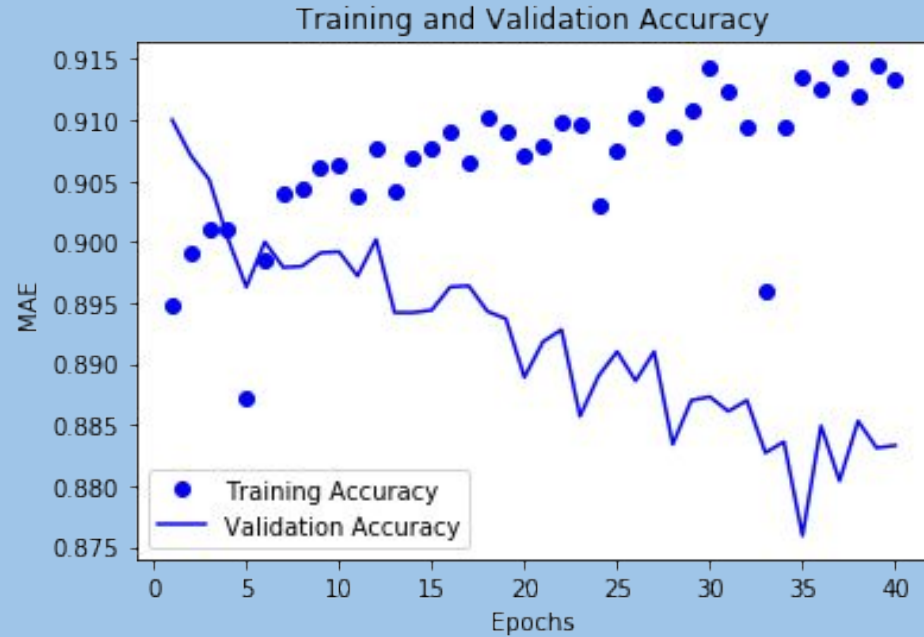
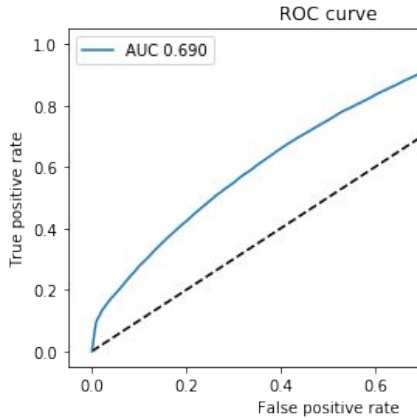
```
[24]: w2v_model.wv.most_similar('brimstone')
```

```
[24]: [('peat', 0.7747483253479004),  
      ('flint', 0.7677335739135742),  
      ('candle', 0.7435697913169861),  
      ('broom', 0.7328553199768066),  
      ('bee', 0.7281134128570557),  
      ('chamomile', 0.7206580638885498),  
      ('wax', 0.719684898853302),  
      ('beeswax', 0.7003276944160461),  
      ('granite', 0.6992579102516174),  
      ('pollen', 0.6967644691467285)]
```

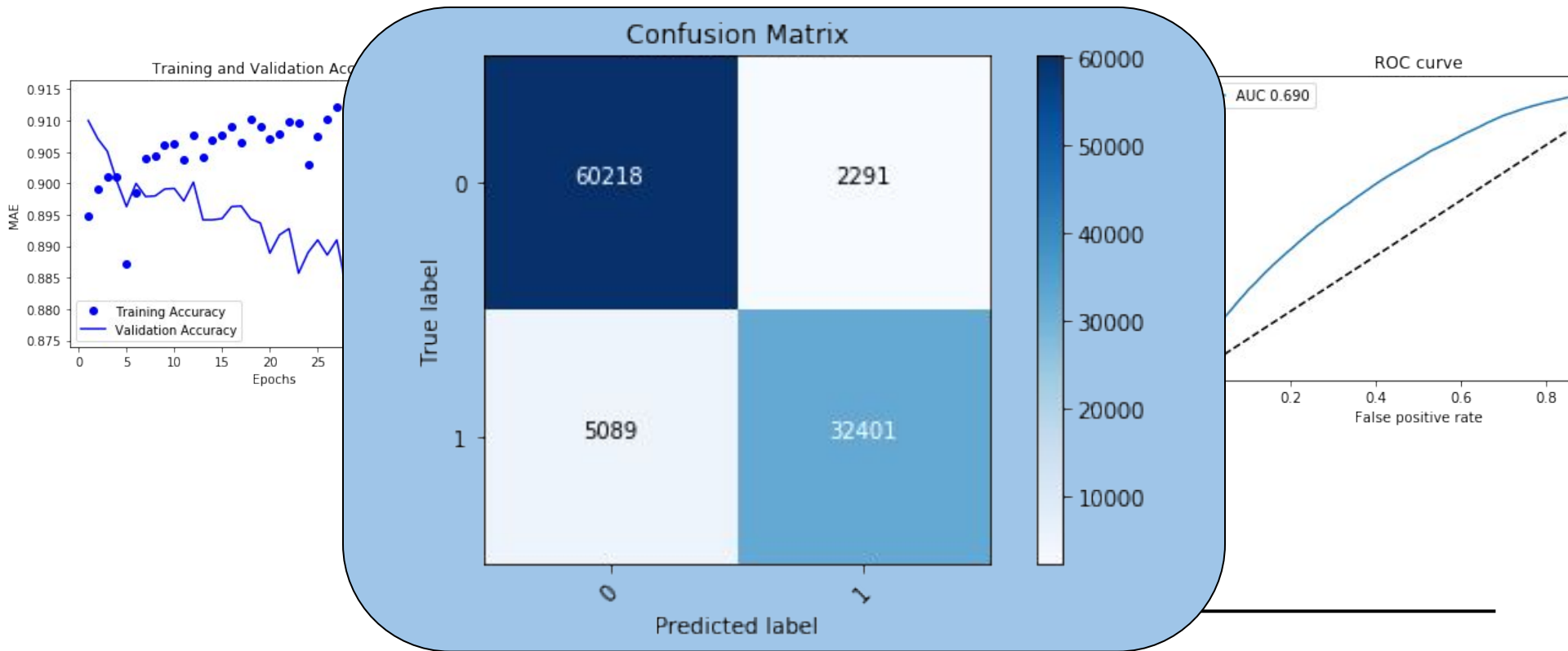
Outcome - Random Forest Model



Outcome - NN Iterative Feedback



Outcome - NN vs Test Data



Suggestions for Further Development

Create data by giving each wine a new name based on their taster description.

Build a live recommender system that pairs likely flavors and suggests wines that fit the search criteria.

Appendix

[Linkedin](#)

[Github Repo](#)

[Blog](#)
