Using Predictive Modelling (SVM) to Predict Winery Sales (Consumption) based on Social Media Data.

GA Data Science - Final Project Outline

James McCutchan

February 13, 2014

1. Problem to be solved

How relevant is a Wineries social media presence to their overall Sales, can their social data be used to predict their yearly consumption?

2. Description of dataset

• Foursquare 'Checkins'

• Facebook 'Page Likes', 'Checkins', 'Talking about this'

• Case consumption per Brand Data from The International Wine and Spirit Research's annual report on consumption of alcoholic drinks

• Using a supervised approach (SVM) and create a decision boundary that accurately separates trivial social media data from non-trivial social media data when plotted against consumption.

• I will train the SVM algorithm by providing it with example data that is grouped into a series of categories. When I provide the algorithm with the test set of data, the model will assign the data to categories based on its resemblance to the know test set.

3. Hypothesis

I hope to gauge whether social media input given to a winery can be used to predict the consumption of that brand.

4. Statistical methods I plan to use and why

This will be a classification problem, finding a decision boundary that can decide whether social data is correlated to consumption

All the data I will be using is labeled so I will not be using clustering techniques (i.e. kmeans). Naive Bayes will not be used because I will need a 'dependent' feature model. SVM will find a 'maximum likelihood' that the social data can predict consumption.

5. Applications the finding may have

I hope this report will provide insight into social media marketing strategies and a visualization of the correlation between social media input and sales.

REFERENCES:

- Foursquare API

- Facebook graph API

- International Wine and Spirit Research's annual report on consumption of alcoholic drinks