Goal
Determine the best model for wine classification in the pursuit of bringing to market a cutting-edge material analysis application which produces accurate, near real-time readings.

Motivation and Objectives

Motivation
Companies have a pressing need to measure material properties throughout their supply chains.

Objectives
- Create an automated data preprocessing system that converts text output files from the portable detectors to CSV files for use in training our models.
- Create models which take the given spectroscopic data and make wine classifications.

Research Challenges
- **Limited Dataset Size**: practical challenges of experimental setup and data acquisition processes.
- **Text Data Format**: requires tailored data wrangling.
- **Abnormalities in the Dataset**: data acquisition process is noisy both at sensor and process levels.

Conclusions
- Both PLS and Artificial Neural Networks are viable methods for the wine classification. More raw data is needed for either model to to be generalized to new unseen data.
- The 5 mm path length led to the highest classification accuracies across both detectors, with the Neural Network Detector 2 - 5 mm candidate model classifying both red and white wines with 100% classification accuracy.
- The PLS Method performed poorly on the red wine data, but showed strong performance on white wine data, as did the Neural Network Method.

Methodology

Data Preprocessing
- Label data, remove calibration runs and data with input errors.
- Convert TXT files to usable CSV files.

Partial Least Squares
- The PLS model binarizes the labels, optimizes the number of PLS components between 1 and 40, and utilizes the number of components that yields the lowest mean squared error.

Neural Network
- Separate data into eight groups, each group representing a candidate model. Two detectors, each with four path lengths at which data was taken.
- Resample data using the bootstrap technique for each candidate model.
- Use the Stratified K Fold strategy to determine training and testing set splits.
- Train a sequential, feed-forward network for each candidate model, using the Keras library with TensorFlow backend.

Results

<table>
<thead>
<tr>
<th>Average Prediction Time (seconds)</th>
<th>White</th>
<th>Red</th>
</tr>
</thead>
<tbody>
<tr>
<td>D1 0.5mm</td>
<td>0.0005</td>
<td>0.0002</td>
</tr>
<tr>
<td>D1 1mm</td>
<td>0.0001</td>
<td>0.0001</td>
</tr>
<tr>
<td>D1 2mm</td>
<td>0.0001</td>
<td>X</td>
</tr>
<tr>
<td>D1 5mm</td>
<td>0.0001</td>
<td>X</td>
</tr>
<tr>
<td>D2 0.5mm</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>D2 1mm</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>D2 2mm</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>D2 5mm</td>
<td>0.0001</td>
<td>X</td>
</tr>
</tbody>
</table>

PLS Classification Accuracy - White Wines

PLS Candidate Model

Detector 1

0.5 mm

1.0 mm

2.0 mm

5.0 mm

Detector 2

0.5 mm

1.0 mm

2.0 mm

5.0 mm

Neural Network Classification Accuracy - White Wines

Neural Network Candidate Model

Detector 1

0.5 mm

1.0 mm

2.0 mm

5.0 mm

Detector 2

0.5 mm

1.0 mm

2.0 mm

5.0 mm

Average Prediction Time (seconds)