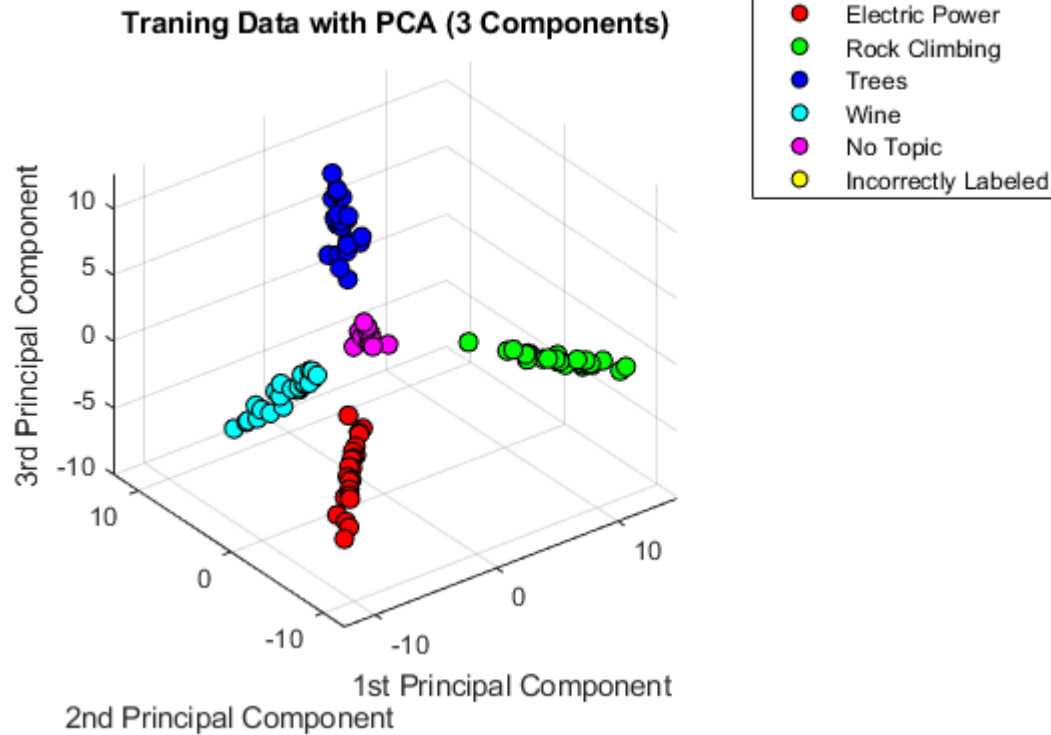
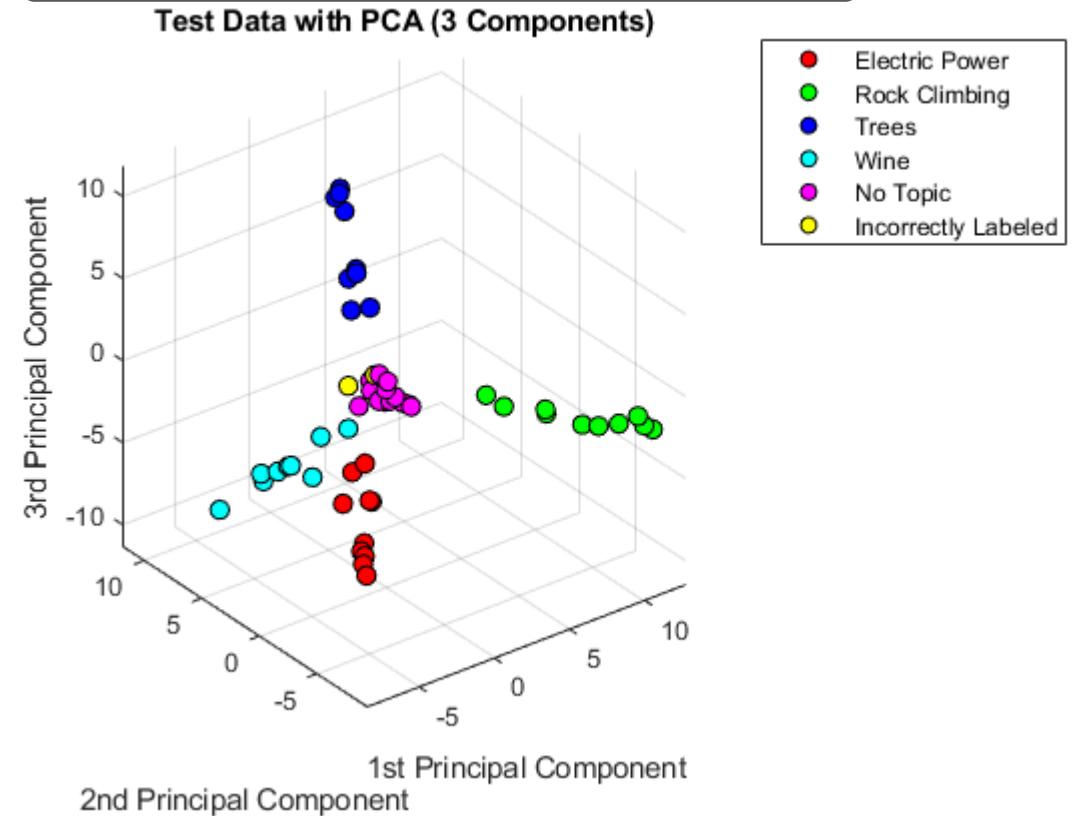


Model Training with Data



Trained Model Applied to New Test Data



Since the data is in a high dimension (number of documents by the number of unique words), the results can only be interpreted by how many documents were clustered in the correct category, and this can be shown with a confusion matrix. However, the above figures use principal component analysis (PCA) to reduce the data by the direction that maximizes the variance of the projected data, preserving as much information as possible as the dimension is reduced. The data can then be visually plotted once the data is reduced to three dimensions.

If you would like to know more about this project, please view the full report [here](#).