1. What is the problem you want to solve

Can we predict when a company will begin exporting their product to other countries based on employee productivity, revenue, space, or other measures? Is there a way to predict which companies are ready to make that transition?

1. Who is your client and why do they care about this problem? In other words, what will your client DO or DECIDE based on your analysis that they wouldn’t have otherwise?

My client is the Economic Development Council of Wisconsin. They want to know which Wisconsin companies may begin exporting in the near future so that they could best leverage grants, marketing, and investment in companies needing help making the transition.

1. What data are you going to use for this? How will you acquire this data?

The data is a collection of several thousand companies with information about number of employees, revenue, what the product is, location, owners, etc. This data has been collected by the Economic Development Council, and provided to me through a data scientist in Wisconsin. Further data that may be needed is to see if companies have begun exporting since this data was collected so that the validity of the model can be compared.

1. In brief, outline your approach to solving this problem (knowing that this might change later).

The basic economic theory is that exporting does not increase productivity, but rather increased productivity leads to exporting. My plan is thus to first, visualize the data and look for simple correlations between things like revenue and number of employees, size, or product. I will isolate the most important variables and extract and clean that data. Then I plan to use logistic regression to make a predictive model for which companies are most likely to begin exporting. I may also explore a decision tree to see if there are any nodes or breaks where companies diverge. Finally, I want to compare this to new data to see how accurate the model is so that it can be applied going forward.

1. What are your deliverables? Typically, this would include code, along with a paper and/or a slide deck.

I will deliver the code and model to the council. Further, a report with several visuals will be created to tell the model’s story and help the council understand the findings.